AD NUMBER
AD484188
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies and their contractors; Critical Technology; MAY 1966. Other requests shall be referred to Rome Air Development Center, ATTN: EMLI, Griffiss AFB, NY 13441-5700.
AUTHORITY
RADC ltr dtd 17 Sep 1971

AD 484188

ADAPTIVE MODULATION AND ERROR CONTROL TECHNIQUES

R. J. Benice
et al

IBM Corporation

[redacted]

FOREWORD

This final report was prepared by IBM Corporation, Federal Systems Division, Rockville, Maryland, under Contract AF30(602)-3603, Project No. 4519, Task No. 451902. The author was assisted by B. W. Beals, F. P. Corr, R. A. D'Antonio, A. H. Frey, Jr., J. Y. Hayase, R. E. Keeler, and J. A. Lewis.

RADC Project Engineer was Mr. Frederick D. Schmandt (EMCRS).

Release of subject report to the general public is prohibited by the Strategic Trade Control Program, Mutual Defense Assistance Control List (revised 6 January 1965) published by the Department of State.

This report has been reviewed and is approved.

Approved: Frederick D. Schmandt
FREDERICK D. SCHMANDT
Project Engineer

Approved: Charles A. Strom, Jr.
CHARLES A. STROM, JR.
Acting Chief
Communications Division

FOR THE COMMANDER: Irving J. Gabelman
IRVING J. GABELMAN
Chief, Advanced Studies Group

ABSTRACT

This report contains the results of an investigation design to develop, analyze, and evaluate techniques to improve the quality of digital data transmission by adapting the modulation or error control parameters to varying channel conditions. Specific systems for varying the bit rate and the alphabet size are postulated and evaluated analytically in realistic channel environments. Variable block length and variable redundancy error control techniques are described and analyzed in a computer simulation of real channels. In addition, the operation and performance of a laboratory prototype of an adaptive error correction decoder that decoder that does not require feedback is described.

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

ILLUSTRATIONS (Continued)

TABLES

Section 1

INTRODUCTION

This report contains the results of an investigation designed to develop, analyze, and evaluate techniques to improve the quality of digital data transmission by adapting the modulation or error control parameters to varying channel conditions. The project was motivated by the observation that fixed parameter techniques currently being employed are designed to match long-term average channel conditions which are, in reality, representative of the actual channel state for only a minor fraction of the time. Intuitively, it appeared that if the system parameters could be matched to various ranges of short-term channel behavior and appropriate parameters used as channel conditions varied, improved performance could be obtained.

Before the system design problems could be considered, two preliminary investigations were necessary. First, the modulation and error control parameters which could be adapted had to be identified and examined for theoretical feasibility and desirability. In addition, in order to ensure meaningful results, a survey of the nature and behavior of real communication channels and their relationships to system parameters had to be conducted. The survey of adaptive techniques is reported in Section 2, and the survey of real channels is reported in Section 3.

The principles and problems of adaptive system design were then investigated. This phase of the investigation, discussed in Section 4, includes mode parameters, criteria for evaluation, delays, control problems, and implementation.

Specific systems for varying the bit rate and the alphabet size were then postulated and evaluated analytically in realistic channel environments. The systems, together with details of the mathematical analysis and the quantitative results. are described in Section 5. In Section 6, variable block length and variable redundancy error control techniques are described and analyzed in a computer simulation of real channels. In addition, the operation and performance of a laboratory prototype of an adaptive error correction decoder that does not require feedback is described.

In general, it was shown that a wide range of adaptive techniques offer significant potential improvement, but that limitations on the reaction time to changes in channel conditions limit the effectiveness of many techniques. The conclusions of the investigation are summarized in Section 7.

Section 2

A SURVEY OF ADAPTIVE TECHNIQUES

The first phase of the adaptive coding techniques investigation was a survey and investigation of the redundancy forms which could be varied to improve performance on a time-varying communication channel. "Redundancy" is used here in its most general sense. For example, adding redundancy could refer to slowing down the data rate, increasing power at the transmitter, employing diversity reception, or to strictly digital techniques like placing additional bits onto the data stream in the format of error correcting codes. The literature reveals very few quantitative results yielding an effective evaluation of tradeoffs among such techniques.

This phase is concerned with selecting appropriate redundancy forms and determining which are theoretically feasible and merit further consideration for use with each combination of modulation technique and channel type. For this last aspect, it was necessary to survey existing channels and modulation techniques to ascertain those combinations which were technically preferable or widely used. This ensured that "real" situations were the subject of the investigation. The survey of channels and modulation techniques associated with them appears in Section 3.

This section concerns identifying the possible redundancy forms to adapt; determining their range of applicability; classifying them with respect to modulation versus coding, where the adapting is performed, when the adapting is performed, and how it is performed; and selecting those techniques which appear to offer the most promise for further investigation.

The investigation was limited to noncoherent FSK, differentially coherent PSK, coherent PSK, and AM suppressed carrier. These modulation techniques cover most cases of current practical interest. Other techniques have been excluded because they are of quite limited use and are either inferior (as in on-off keyed AM) or are experimental and have insufficient data available (as for RAKE). These exclusions did not restrict the applicability or significance of the results.

2.1 TYPES OF VARIABLE REDUNDANCY TECHNIQUES

A considerable number of potentially variable redundancy forms were selected for investigation. These include variable information content per symbol (higher-order alphabets), variable decision mechanism for incoming signals (variable decision threshold, space diversity, frequency diversity, time diversity, polarization diversity, angle diversity), variable symbol length, variable power, and variable error control (variable redundancy, variable block length, variable use of redundancy, variable number of transmissions per message).

While this list does not include every conceivable adaptive technique, it does include those which have received considerable attention. The results derived for these techniques, which are representative of the various adaptable aspects of digital data transmission, should have analogues for other techniques that might be considered at some future time. These techniques have been characterized, examined for their applicability for the various combinations of channels and modulation techniques, classified with respect to their application points, surveyed for dependence on or independence from feedback, and classified with respect to the effect of their utilization on throughput and error rate.

2.1.1 Variable Modulation/Demodulation Techniques

2.1.1.1 Variable Bit Length

The error rate in both coherent and non-coherent reception is a decreasing function of E/N_0 where E represents signal energy per symbol and N_0 is the noise energy (i.e., watts per cycle per second). The value of E will fluctuate as the received signal fades. The two ways to compensate directly for any reduction in E are to increase transmitted power or symbol length.

Varying the length (i.e., duration) of a transmitted bit provides a means for varying the energy per bit as the signal level or noise level fluctuates. Given proper synchronization and control, there is no reason that such an adaptive system could not be used for all the combinations of channel type and modulation technique under consideration.

In conjunction with retransmission systems, varying the bit length was shown to be superior to fixed block coding and variable block length coding for both independent error channels and Rayleigh fading channels by Corr and Frey [1]. The problem has also been investigated for fading channels (with ideal feedback links) by Lieberman [2]. He also compares this technique with varying the transmitter power (see subsection 2.1.1.3).

In operation, a variable bit length system would tend to preserve a particular error rate at the expense of a variable throughput. The operational aspects of this technique are investigated in more detail in Section 4 and its performance is analyzed in Section 5.

2.1.1.2 Variable Bits per Symbol (Higher-Order Alphabets)

It is quite common to use multiple-level signalling alphabets in which each transmitted symbol conveys more than one bit of information. In multiple frequency shift keying (MFSK), an m-ary alphabet is achieved via a set of m separate, spaced tones and with phase shift keying (PSK) one of m possible phases is used. Adaptive systems using higher-order alphabets can improve communications in two ways. By halving the symbol rate and doubling the order of the alphabet (i.e., the number of bits per symbol) the data rate is maintained and the error rate for a given signal power-to-noise power density ratio is reduced. On the other hand, just halving the order of the alphabet reduces error rate at the expense of lower data rate.

The effectiveness of this technique will depend on whether the dynamic range of the signal power variation due to changing channel conditions can be successfully accommodated by adapting the number of phases per pulse or the number of separate frequencies. The use of a large range of values for m will be required to accommodate a large received signal power dynamic range.

Adaptive systems that would vary the number of levels, m, are theoretically possible for all of the channel and modulation combinations under consideration, provided the control and synchronization requirements can be met. Fixed higher-order techniques have been investigated extensively. Equations and curves for error probabilities in various environments have been published by Reiger [3], Sussman [4], and Turin [5]. These investigators, among others, have compared the performance of the different alphabet sizes but have not investigated changing the number of levels while the system is operating.

Operational aspects of varying the alphabet size are considered in Section 4; the performance is quantitatively analyzed in Section 5.

2.1.1.3 Variable Transmitter Power

Varying the transmitter power offers a means of varying the energy per signalling element to compensate for variations in the signal-to-noise ratio.

Lieberman [2] has investigated variable power together with variable bit length for a slowly varying medium. Axelby and Osborne [6] have also considered the problem. Varying the power would tend to preserve a given error rate while operating at a fixed throughput rate.

This technique appears to be, at best, of limited value. In most systems, power is peak limited rather than average limited. However, advocates of variable power claim savings in money and reduction of RFI problems.

On land lines a transmitter power limitation is imposed upon the data transmission user by the common carrier. The main purpose of this limitation is usually to prevent cross-talk between channels using carrier systems. The input power limitation is relatively low—typically between 15 db below a milliwatt and a milliwatt. Although increasing power level reduces error rate, an adaptive power level adjustment system for land lines is not considered useful. Since the maximum allowable level is already low, the transmitter should be continuously set at this level. The cost saving involved in reducing this power level during good transmission periods is minuscule at best.

For microwave line-of-sight links, adaptively variable power level still does not yield substantial cost savings because output power is also low in such systems (e.g., 1 watt). Again, the proper technique is to transmit at a level high enough to achieve satisfactory transmission nearly all the time. During periods of good transmission the cost reductions (or even RFI reduction) achieved by adaptively reducing output power are not significant.

2.1.1.4 Diversity Techniques

The various diversity techniques are adaptive techniques which are currently being used over many links. They differ from most of the other techniques in that only the receiver adapts. Redundancy is introduced into the system in the form of multiple antennas or receivers which provide multiple inputs to the decision mechanism at the receiving terminal. The signal which is accepted varies. Scanning diversity techniques select the first signal encountered which meets predetermined criteria; optimal selection diversity selects the best from all of the received signals; combining diversity techniques use all the received signals.

In spaced antenna diversity, more than one antenna is used in various locations defined with respect to the great circle passing through the transmitter and receiver. The different antennas provide independent versions of the same signal. There is also a degree of independence between signals arriving at a given location at different angles. Angle diversity uses narrow beam width antennas to obtain these different versions. In some cases (e.g., ionospheric reflection) the transmitted signals undergo changes in polarization. It has been observed that the horizontal and vertical components fade independently. Consequently, polarization diversity can be used, where two dipoles (polarized horizontally and vertically) obtain the independent copies of the incoming signal.

There are two diversity techniques in which the transmitter plays a major role: time diversity and frequency diversity. In time diversity, identical signals are sent spaced in time to enable correlation techniques at the receiver to determine the true signal. Since fading varies with time and becomes independent as the spacing increases, this technique can be quite successful. However, the repetition of information drastically reduces the throughput rate. In frequency diversity, the signal is sent over several distinct frequencies. Since fading varies with frequency, several versions of the same signal are produced at the receiver. This technique, however, makes inefficient use of bandwidth.

Lawton [7] investigated the effect of diversity on signal-to-noise ratios and found that, for an optimum coherent diversity combiner, the ratio of the probability density functions without diversity and with m-fold diversity is given by

$$\frac{P\,(W \mid \text{with no diversity})}{P\,(W \mid \text{with m-fold diversity})} = \frac{1}{(m-1)!}\left(\frac{W}{R_0}\right)^{m-1} \qquad (2\text{-}1)$$

where W is the output signal-to-noise ratio and R_0 is the mean signal-to-noise ratio at each input. It was assumed that all inputs have independently distributed Rayleigh fading statistics with equal means.

The relationships between the system error rate, the order of diversity, the received average power during a pulse on-time, and the average energy per transmitted bit to noise power density ratio were also analyzed by D. P. Harris [8].

Diversity techniques are of considerable value in media subject to fast fading. They are not applicable to land lines, satellites, or line-of-sight VHF which are not significantly affected by fast fading.

In practice, diversity techniques would be used in combination with the other adaptive techniques being considered. Diversity techniques would overcome the effects of fading which is too fast to adapt to by the other means.

For long-term fading, diversity reception is of no use as an adaptive mechanism, since the slow fades are characteristically flat fades affecting all frequencies of interest equally. Diversity techniques are fixed throughput techniques which improve, but do not fix, the error rate.

2.1.1.5 Null-Zone Techniques

Null-zone techniques, as used here, refers to those techniques in which the receiver need not make a decision as to whether a particular received bit was a "mark" or a "space." Instead, part of the usual mark and space decision region is designated a no-decision region in which signals are not decided upon. With feedback, received signals falling into the null-zone are left un-decided and a retransmission is requested. Without feedback, received signals falling into the null-zone are considered erasures and are corrected by so-called erasure-correcting codes which are similar to forward error-correcting codes. Thus null-zone techniques can be fixed or variable throughput systems.

The use of null-zone techniques, both fixed and with the boundaries of the null-zone variable, has been extensively investigated by Harris et al [9].

These techniques have not gained widespread acceptance. Indeed, it is not clear that they offer any gain over soundly designed error-detection-retransmission systems or powerful long-block forward error correcting codes. In theory, they are applicable to all of the channel-modulation combinations under consideration.

2.1.2 Variable Error Control Techniques

In addition to the various adaptive techniques which vary the redundancy by varying modulation parameters discussed in the preceding sections, there are numerous techniques which adapt by varying some aspect of error control coding.

Error control techniques can be varied in many ways. The redundancy, the block length, the mode, the amount and type of correction, and the number of transmissions per message can all be varied. Error control techniques have an important place in the design of reliable digital data transmission systems since systems with optimally chosen parameters may still not give a satisfactory error rate. Variable error control techniques can be designed for constant throughput with improved error rate or for constant error rate with improved but variable throughput.

2.1.2.1 Variable Block Length with Fixed Redundancy

Varying the encoded block length while keeping the number or redundant bits per block fixed affords an opportunity to increase the data throughput rate in periods of good channel behavior and to slow down when conditions deteriorate. The fixed amount of redundancy can provide error correction at all operating lengths with decreasing capability as block length increases. Since the number of redundant bits remains constant, the code polynomial and hence the encoder and decoder can be the same, except for counters, in all modes. Although service bits could be used by the transmitter to indicate block length, it appears that this technique will be most useful when feedback is present.

This technique can be used for an error detection-retransmission system or for a forward error correcting system.

In the case of an error detection-retransmission system, the Adaptive Dual-RQ logic is well suited for controlling the system. This logic was developed by Corr and Frey [1]. They also designed, analyzed, and determined optimum block lengths for variable block length retransmission systems. This technique enables the power of the detection-retransmission technique to be extended. In periods of low error rate, long blocks with very high percentages of data can be transmitted and still be protected by an extremely reliable error detecting code. In periods of high error rate, the reliability of the code is maintained while the number of bits to be retransmitted as a result of the detection of errors is decreased. It has been demonstrated that for any given degree of error protection the optimum (with respect to throughput) block length increased with decreasing error probability [10].

In the case of a forward error correction system, control information from the receiving terminal to the transmitting terminal can be conveyed via the new logic described in subsection 4.6.2.2. This logic seems particularly well suited to non-retransmission adaptive systems. Through skillful choice of the fixed code polynomial, the system can offer the strong error correction of high percentage redundancy in higher error rate periods and less protection for the high data rate transmissions during lower error rate periods. This technique can be used with independent error correction or burst error correction.

Monitoring the received signal level via the error correcting code already in the blocks to determine the current error rate was used in the real-channel simulation of the forward error correction variable block length technique. A further discussion, including quantitative results, appears in Section 6. Hardware considerations for implementing this technique are discussed in subsection 4.7.

2.1.2.2 Variable Redundancy in a Fixed Block Length

Varying the number of redundant bits in a block of fixed length is another way of increasing and decreasing throughput with corresponding decreases and increases in error protection. This technique is simpler with respect to block synchronization but requires different coding polynomials for each mode. Chien and Tang [11] have developed several types of variable redundancy codes.

As was the case with the variable block length techniques of the preceding section, variable redundancy can be used for either forward error correction or detection-retransmission. Channel quality can be monitored by the same means as for variable block length techniques.

When used as a retransmission system the regular Dual-RQ logic rather than the Adaptive Dual-RQ can be used for control. The fixing of the block length results in a simpler system than the variable block length system. There does remain, however, a problem of alignment in the variable redundancy system corresponding to the problem of the variable block length system. This is discussed in subsection 4.6.1.2.

Variable redundancy, fixed block length techniques can also be used for forward error correction. There are two possible control systems for this case. The first system uses the new logic developed for the variable block length, fixed redundancy, technique. Another possibility is to use two decoders and use a polynomial framing technique to recognize the polynomial, and hence the code, being used [12]. In addition, it appears that there is no reason why the variable redundancy technique cannot be extended to include interleaved block codes or recurrent codes. Implementation of this technique is considered in subsection 4.7. The results of the computer simulation and evaluation are presented in Section 6.

2.1.2.3 Variable Utilization of Fixed Redundancy

In addition to the techniques of the two preceding sections, it is also possible to keep both the block length and the redundancy (and hence the throughput rate) fixed but to vary the interpretation and utilization of that redundancy in order to improve the error rate. This technique would be valuable in channels subject to different types of errors. Suppose that a channel is characterized by burst errors but that independent errors occasionally hit encoded message blocks. Using a special technique (the implementation for which is discussed in subsections 4.6 and 4.7), many polynomial codes can be used for either independent error correction or burst error correction, depending on the decoding algorithm. It is possible to design a system capable of employing both types of decoding and switch between them as necessary. The only adaptive changes that are made when using this technique are made in the decoding equipment at the receiving terminal. The state of the channel is also detected at the receiving terminal, usually in the decoder.

This leads to several advantages of this technique over other adaptive techniques. Since no modifications are made in the encoder, there is no need to transmit information back to the transmitting terminal. Thus, this technique can be applied in systems where no feedback channel is available or where it is difficult to insert protected feedback information on a return path. Since the equipment for sensing the state of the channel and the equipment to be adapted to the channel state are co-located, there is no delay required for updating information to be transmitted. Thus, this technique has a quicker reaction time than techniques requiring updating information to be passed between the terminals. Since it is the inability to react quickly to changes in the channel that causes the downfall of many adaptive techniques which appear abstractly to have merit, the fast reaction capability of this technique makes it particularly attractive.

Unfortunately, this technique cannot be analyzed fairly in any model or simplified channel. Indeed, the assumption of either a random error channel or a burst error channel would render the adaptive feature useless. On the other hand, the construction of an artificial "sometimes random, sometimes burst" channel would ensure that this technique would give superior performance. Unlike the other techniques, its performance is highly sensitive to real channel idiosyncrasies, and, consequently, can only be evaluated against real channel serial bit error characteristics.

Fortunately, a working feasibility model of such a system has been built and was available for use. The limited number of real channel tapes made available were run through the adaptive decoding device in the laboratory. The results of these tests are described in Section 6.

2.1.2.4 Variable Error Control Mode

Variable error control mode systems offer an interesting new possibility which does not appear to have been investigated. Such techniques, if feasible, would switch between error detection, forward error correction, and error detection with retransmission depending on the variations in channel conditions.

For example, such a system might use its correction capability when small numbers of random errors or short bursts occur and use its retransmission capability for heavier concentration of errors or for longer bursts. The regions of superiority for these two modes were charted by Benice and Frey [10] . As another example, a retransmission system might switch to a detection only system in error periods of great length in order to avoid overtaxing the storage at the transmitter.

The usefulness of this technique depends considerably on the user's needs and requirements as well as on fine-grained descriptions of real channels. Furthermore, the number of possible combinations of modes is staggering. In view of these complexities, plus the complexity of the implementation and control problems, this technique has not been analyzed here.

2.1.2.5 Retransmission Systems

Retransmission systems vary the redundancy by varying the number of times that a message is transmitted. In effect, correct messages employ only enough redundancy to ensure that they are properly identified as being correct while messages received in error are repeated until they are received correctly.

Here, discarded transmissions are classified as redundancy. Thus, the throughput rate goes up in error-free periods and down in periods in which errors occur. This technique has already been exhaustively analyzed and compared with other techniques by Benice and Frey [10, 13].

Several retransmission systems with simpler implementation than those previously considered have been developed. These are essentially the same as Dual-RQ but differ in the nature of the control message. Their matrices of transition probabilities have been found, and a program has been run to compute their throughput and error rate and to compare them with the original Dual-RQ logic. The results are described in subsection 4.6.1.3.

2.2 LOCATION OF THE ADAPTIVE PROCESS

The various adaptive techniques under investigation vary considerably in the place at which the adaptive procedure is actually performed. For example, varying the peak power per symbol is performed at the transmitting terminal with no adaptive action required at the receiving terminal. On the other hand, with the diversity techniques and the null-zone techniques, the adaptive action is performed ~~solely~~ at the receiving terminal. Similarly, the variable utilization of fixed redundancy and block length is a receiver-only adaptive process.

Most of the advanced techniques which received attention in this study required both the transmitting and receiving terminals to vary their operating mode and to coordinate these changes with each other. These techniques are the least understood and present the most practical problems. The system design problems for the four most promising techniques in this category—variable block length, variable redundancy, variable bit rate, and variable bits per symbol—are attacked in Section 4.

2.3 FEEDBACK REQUIREMENTS

Some of the variable redundancy techniques under consideration are designed for one-way channels; some require a feedback link, others can be designed to operate either way.

Diversity techniques generally do not employ feedback. The switching or combining process at the receiving terminal effects the variation in such techniques. Variable power techniques generally require feedback, but some cases (e.g., satellites with predictable range variation) can be handled independently by the transmitting terminal.

There are variable bit length, variable bits per symbol, and variable decision threshold techniques that are adjusted at the transmitter without feedback, but the techniques that are most effective and most responsive to shorter term variations are those in which the receiver monitors the channel and continuously conveys control information.

Error control techniques exhibit the widest variation with respect to dependence and independence from feedback. Numerous forward acting techniques operate without feedback, whereas retransmission techniques require it. Variable mode systems can operate with and without feedback. Variable block length and variable redundancy systems can operate without feedback but cannot be as responsive. The most effective techniques are those with continuous monitoring and adapting. The variable utilization of fixed redundancy technique is exceptional in this regard, in that it adapts quickly to changes in the channel without feedback. It is, however, a fixed throughput system and cannot take advantage of low error rate intervals by significantly stepping up the throughput rate.

Section 3

A SURVEY OF REAL CHANNELS

To ensure meaningful results, a survey of the nature and behavior of real communication channels was conducted. For the various channel types commonly used, the modulation techniques generally associated with them, either for reasons of technical superiority, tradition, or availability, were ascertained. Thus non-optimum techniques were admitted, for it was the intent of this study to determine means for improving wide classes of existing channels—not to suggest changes in the modulation techniques or media used.

Some modulation techniques were excluded because they are of limited use and are either generally inferior (e.g., on-off keyed AM) or experimental and have insufficient data available for our purposes (e.g., RAKE). This exclusion does not restrict the applicability of results on adaptive data transmission techniques.

The following channels were selected for consideration: satellites, land lines and submarine cables, ionospheric scatter, ionospheric reflection, tropospheric scatter, line-of-sight microwave, and knife-edge diffraction. Although this list does not contain all channels that possibly could be used, a wide variety of channels, including those used to a significant degree at the present time, are represented. Solutions to their problems will have applicability to the problems of other channels. The lack of measured performance data for other types of channels would immediately block any attempt to analyze the performance of adaptive techniques.

For each of the channel-modulation combinations, the literature was searched for data on fading and error characteristics. Since data of this nature was inadequate for analysis and simulation, models of channel behavior were needed. The collected data provided values for the parameters in the models and ensured their validity. The models could then be used to provide continuous long-term descriptions of the channels which were consistent with actual measured behavior.

3.1 MODELS OF CHANNEL BEHAVIOR

3.1.1 Rayleigh Model

When a pure tone is emitted into a medium by a transmitter, the signal received at a remote location is, in practice, more complex than the simple transmitted sinusoid. One mathematical model used to describe this incoming signal for several media is the Rayleigh fading model. This model is used for several reasons: it is physically reasonable, can be empirically verified, and is mathematically tractable.

The Rayleigh probability density function for the envelope E of the received wave is given by

$$f(E) = \begin{cases} \dfrac{2E}{\overline{E^2}} \exp\left(\dfrac{-E^2}{\overline{E^2}}\right), & \text{for } E \geq 0; \\ 0, & \text{for } E < 0 \end{cases} \tag{3-1}$$

where $\overline{E^2}$ is the time-average of E^2. The probability distribution function is then given by

$$F(E) = \int_0^E f(E^*)dE^*$$

$$= 1 - \exp\left(\frac{-E^2}{\overline{E^2}}\right) \tag{3-2}$$

The density function for the received signal power P is then

$$f(P) = \frac{1}{\overline{P}} \exp\left(\frac{-P}{\overline{P}}\right) \tag{3-3}$$

where $\overline{P}$ is the time-average of P, $P = S^2 = E^2/2$, and S is the short term average received RMS signal level.

From a physical point of view, Rayleigh fading would occur for a medium with the following properties:

1. The received signal is the sum of signals arriving on many different paths from the transmitter.
2. The strength of the signal on each path is constant, or nearly so.
3. There is no path (or paths) which dominates in the sense that its received strength is an appreciable portion of the total power received.
4. The individual path lengths vary with time.

The reason that this combination of events produces Rayleigh fading is similar to that often used by statisticians to justify the use of a gaussian hypothesis, namely the central-limit theorem. This theorem states that when a large number of numerical values are taken from different (or the same) sample spaces and added, the distribution of the sum tends toward a normal or gaussian law under very broad conditions. In the fading case, one further step is required—relating the naturally occurring normal deviate to the "envelope" of the received signal.

The signal emitted is of the form:

$$s_c(t) = A \cos(\omega_c t + \theta) \tag{3-4}$$

where A is the amplitude, ω_c is the frequency and θ is the phase of pure tone emitted.

A signal received over a multipath medium will be of the form:

$$s_r(t) = \sum_{i=1}^{n} k_i(t) A \cos\left(\omega_c t + \psi_i(t)\right) \tag{3-5}$$

where $k_i(t)$ is the attenuation of the i^{th} path, $\psi_i(t)$ is the phase of the signal received on the i^{th} path, and n is the number of discrete paths.

Equation (3-5) may be related in a straightforward manner to the four properties mentioned above for a Rayleigh fading medium. Property 1 implies that n is large. Property 2 implies that the $k_i(t)$ are constant or slowly-varying functions of time. Property 3 implies that no k_i is substantially greater than all the others. Property 4 implies that the ψ_i are not constant.

Equation (3-5), within very broad restrictions, can be written in the following form:

$$s_r(t) = E(t) \cos\left(\omega_c t + \gamma(t)\right) \tag{3-6}$$

where E(t) is the amplitude of the composite signal and $\gamma(t)$ is the phase of the carrier of the composite signal.

Equation (3-6) indicates simply that a wave consisting of a sum of sinusoids with individually varying phases and amplitudes can be considered as a single composite wave. This is simply phasor addition. The amplitude of the composite signal, E(t) in Equation (3-6) is the received envelope. If the incoming signal were passed through a conventional AM detector, $E(t)$ would be observed. Further, it is $E(t)$ which will exhibit the Rayleigh distribution if the medium has the four properties described. Why the central-limit theorem leads to such a conclusion is discussed below.

An exercise in trigonometry yields the alternative form to Equation (3-6):

$$s_r(t) = x(t) \cos \omega_c t + y(t) \sin \omega_c t \tag{3-7}$$

where

$$[E(t)]^2 = [x(t)]^2 + [y(t)]^2 \tag{3-8}$$

$$\gamma(t) = \tan^{-1} [y(t)/x(t)] \tag{3-9}$$

If x and y are random variables with gaussian distributions, then E, the square root of $x^2 + y^2$, will be distributed according to the Rayleigh law. This distribution is equivalent to what statisticians call a chi distribution with two degrees of freedom.

Returning to Equation (3-5), we can determine $x(t)$ and $y(t)$ by multiplying the right side of the equation by $\cos \omega_c t$ and $\sin \omega_c t$ and retaining only those components without a $2\omega_c$ term. After performing this operation, it becomes clear that x and y are, indeed, the sums of a large number of random variables. If the central-limit theorem is applicable, x and y are normally distributed and, by definition, E is Rayleigh-distributed. The four properties stated are roughly those for which the central-limit theorem should apply.

The conditions described above are characteristic of scatter media. Indeed, it has been found (see subsection 3.2) that the model is valid for short term fading on tropospheric scatter and ionospheric scatter channels. For such channels one need know only $\overline{E^2}$ in (3-1) to define the distribution. Unfortunately in many cases slow fading must also be considered and the distribution of $\overline{E^2}$ will be required (subsection 3.1.4).

3.1.2 Rician Model

Various media display characteristics different from those leading to Rayleigh fading. On certain media (e.g., HF via ionospheric reflection), one path predominates. That is, the received signal can be thought of as a single strong discrete component of 1 volt RMS normalized plus a mixture of signals which would combine to form a Rayleigh envelope in the absence of the discrete term. In this case, the so-called Rician distribution is given by

$$F(E) = \frac{2}{\psi_0} \int_0^E E^* \exp\left[-(1+E^{*2})/\psi_0\right] I_0\,(2E^*/\psi_0)dE^* \qquad (3\text{-}10)$$

where ψ_0 is the mean-squared value of the Rayleigh components and $I_0(z)$ is the modified Bessel function of the first kind. The corresponding probability density function is given by

$$f(E) = \begin{cases} \dfrac{2E}{\psi_0}\, I_0\,(2E/\psi_0)\, \exp\left[-(1+E^2)/\psi_0\right], \text{ for } E \geq 0; \\ \\ 0, \text{ for } E < 0 \end{cases} \qquad (3\text{-}11)$$

For a value of the ratio of the amplitude of the specular component to the square root of the average power from other sources greater than about 3, the distribution behaves like the gaussian, with mean value of the envelope equal to the amplitude of the specular component, and variance equal to the total average power for the noncoherent paths [14, 15]. Thus, the distribution density of the envelope can be approximated as

$$f(E) = \frac{1}{\sqrt{2\pi\,\psi_0}} \exp\left(\frac{-1}{2\,\psi_0}(E-\bar{E})^2\right) \tag{3-12}$$

where $\bar{E}$ is the amplitude of the dominant component. The corresponding density function for the received signal power is given by

$$f(P) = \frac{1}{\sqrt{2\pi\,\psi_0\,P}} \left\{ \exp\left[\frac{-1}{2\psi_0}\left(\sqrt{2P} + \sqrt{2\bar{P}}\right)^2\right] + \exp\left[\frac{-1}{2\psi_0}\left(\sqrt{2P} - \sqrt{2\bar{P}}\right)^2\right] \right\} \tag{3-13}$$

Thus the Rician model (or the gaussian approximation) needs an additional input parameter in order to be used, namely, the ratio of the power in the discrete sinusoid to the power in the Rayleigh components. Methods for determining this ratio were considered by McNichol [16] and Norton et al [17].

3.1.3 Other Variations from Rayleigh Fading

It was shown that Rician fading resulted when property 1 of the Rayleigh model was not satisfied. Other deviations from the Rayleigh occur if any of the other three properties are violated. For example, if the number of discrete paths is only two or three, property 1 does not exist. In this case, the central limit cannot take effect and cause x and y to be gaussian. Moreover, even when the number of paths is moderately large, the tails of the x and y distributions will not be gaussian. The resulting deviation from Rayleigh for the envelope at the distribution tails has often been observed empirically.

If property 4 were violated, no fading would occur since the sum of a group of phasors of the same frequency and constant phase is a resultant phasor of fixed phase. In practice, media are never perfectly phase-stable, and an interference between paths occurs. The envelope at any instant depends upon the relative phasing.

Thus, Rayleigh and Rician fading are the result of multipath propagation with changing path length. This is the so-called rapid fading. Should the strength of individual paths change, the mean-squared value of the Rayleigh distribution will change. Fades of this nature are referred to as slow fading and are discussed in subsection 3.1.4.

Some experimenters have found deviations from the Rayleigh distribution on several media. Differences from these distributions, especially in the tails, have been noted on empirical curves. There are several possible explanations for these deviations:

1. The Rayleigh or Rician distribution is approached only as the number of discrete paths grows large.
2. Unknown local conditions can have unusual effects.
3. The amount of data in the tail region may be insufficient.
4. The measuring technique may not have suppressed the effects of long-term fading.

In conclusion, the short-term Rayleigh (or Rician) fading model has been found to be acceptable for this study on the grounds that (1) it agrees with most empirical measurements on the media examined, (2) it is mathematically tractable, and (3) several related distributions and functions have been tabulated.

3.1.4 Slow Fading

While the Rayleigh and Rician models appear to be valid for short-term fading (i.e., minutes or less), they cannot account for the slower variations in the median received power level.

The long-term distribution has been treated as log-normal by Bullington et al [18] and several other authors. Typically the empirical distribution of the received power in db is plotted on normal probability paper. The result approximates a straight line leading to the conclusion of a log-normal distribution for the short-term median or average of the received power.

Siddiqui and Weiss [19] proposed a two-parameter gamma distribution in place of the log-normal for the short-term median of received power. Although Siddiqui and Weiss admit to a lack of sufficient empirical data for an adequate log-normal versus gamma comparison, they do point out that the gamma is analytically tractable as a natural outgrowth of the Rayleigh. Another problem with both gamma and log-normal is that neither takes account of seasonal and diurnal variations since they both represent stationary time series.

The discussion to this point refers only to the amplitude distribution of the long- and short-term fades. Complete statistical descriptions would require multidimensional distributions. The paucity of measured data of this nature would make it an almost hopeless task to verify any model rigorously.

For our purposes we will use the log-normal model. It has wide acceptance, is analytically tractable, and is easily related to available measurements of channel statistics. It will be necessary to account for diurnal and seasonal variations separately. The autocorrelation function for the median received power in db is sufficient to define all the required statistics.

3.1.5 Gilbert Model

Some channels (e.g., land lines) are not subject to a slow variation in signal strength, but instead are subject to impulse noise and dropouts which cannot be covered by any of the proceding models.

The impulse noise telephone channels are generally characterized by the Gilbert model [29]. This model has an error-free good state and a bad state in which errors occur with probability 1-h. The transition probabilities from bad to good and good to bad are denoted p and P, respectively. This model has proved to be quite satisfactory for telephone channels, although not for fading channels. Several refinements have been developed which use more than one bad state or allow infrequent (about 1 in 10^5 or 10^6) errors in the good state.

Extensive measurements have been made for various telephone lines. From the Alexander, Gryb, and Nast investigation [21], values in the ranges 0.8 to 0.9 for h, 0.05 to 0.10 for p, and 10^{-4} to 10^{-3} for P appear to be typical. Recently reported work at Mitre [22] indicated $h = 0.56$, $p = 0.14$, $P = 0.169 \times 10^{-5}$ for their link.

3.1.6 Pareto Model

Berger and Mandelbrot [23] and Sussman [24] considered a channel model in which the distribution of the intervals between errors is described by a Pareto law. Specifically, if the numbers t_i represent the positions of earlier errors, and the random variable $T_{n+1} - t_n$ represents the distance between successive errors, then the distribution

$$F(t) = \text{Prob}\ (T_{n+1} - t_n < t) \tag{3-14}$$

is statistically independent of the t_i and in general satisfies

$$\begin{aligned} 1 - F(t) &= t^{-\alpha},\ \alpha > 0, \text{ for small } t \\ 1 - F(t) &= Pt^{-\alpha'},\ \alpha' > 0, \text{ for large } t \end{aligned} \tag{3-15}$$

This results in higher probabilities for small and large t than would result from a geometric distribution and lower probabilities for medium values of t. For simplicity, it was assumed that $\alpha = \alpha'$ (whence $P = 1$).

Comparisons with measurements made on real telephone circuits indicate that this model gives a better fit than the Gilbert model. However, the Pareto model is more difficult to use, and its parameters are more difficult to determine. Like the Gilbert model, the Pareto model does not appear to be applicable to the fading channels.

3.1.7 Independent Error Model

Another widely used model assumes that errors occur independently with some fixed probability p. The main virtues of this model are the ease with which values of p can be ascertained for real channels and the ease with which the model can be used, either analytically or in a computer simulation. The main drawback of the independent error model is its lack of validity due to the lack of agreement between this model and the clustering of errors observed on almost all channels. There is evidence, however, that some satellite links are characterized by independent errors [25, 26].

3.2 MEASURED CHANNEL BEHAVIOR

Parallel to the analytical considerations, the substantial amount of data which has been accumulated in the literature from numerous measurement programs is useful in understanding the behavior of the envelope function E(t). To understand the significance of this data, three characteristic time intervals used by experimenters and analysts to study the statistical properties of the envelope can be distinguished.

Using Baghdady's notation [27] , the shortest duration of time of interest is T_{short}. During this time the signal level remains substantially constant. In the case of a carrier modulated by some information-bearing signal, the carrier component does not fluctuate significantly during T_{short}. This time interval is some integral multiple of the basic information bit duration. One can then define average signal-power-to-noise-power ratio as

$$r(t) = \frac{S^2(t)}{N^2(t)} = \frac{\frac{1}{T_{short}} \int_{t-T_{short}/2}^{t+T_{short}/2} s^2(t)\ dt}{\frac{1}{T_{short}} \int_{t-T_{short}/2}^{t+T_{short}/2} n^2(t)\ dt} \qquad (3\text{-}16)$$

when $f(t) = s(t) + n(t)$ is the signal received. Typical values of T_{short} are less than a few milliseconds.

To study the fast fading phenomena, experimenters use T_{int}, a longer time interval than T_{short}. During this time interval, significant changes are observable in the carrier component. However, T_{int} is not long enough for observing a change in the mean or median value of the carrier component. The duration of T_{int} is dependent upon the mode of propagation, frequency, distance, etc., but investigators have typically used values of a few seconds or minutes.

The third duration of interest is T_{long} in which variations in the median or mean carrier component can be observed. To assess the slow fading phenomena, readings T_{long} apart are taken. Typically, T_{long} ranges from a few minutes to several hours. The distribution of the hourly depth of fade, which is the ratio of signal levels exceeded for 50 percent and 99 percent of each hour expressed in db, is a useful channel parameter in characterizing fading conditions over a long period of time.

The concept of hourly depth of fade does not provide any information concerning how many fades occur per unit time or the duration of the fast fades. The level of the fade can be expressed in terms of the signal level below the hourly median signal value. With this quantity as a parameter, the number of fades per hour can be plotted in terms of the fade duration.

Thus, fading can be described in terms of the extent of fading, duration of fade, and the number of fades that occur in a given duration. The extent of fading can be further expressed in terms of parameters derived from measured data, e.g., hourly depth of fade, signal level below the hourly median signal level, fading range, and system loss. The fading range is the ratio of the signal level exceeded 10 percent of the hour to the signal level exceeded 90 percent of the hour. The system loss is defined as the ratio of the power delivered to the transmitting antenna to the power output of the receiving antenna, exclusive of transmission line losses. This parameter provides a measure of the absolute power loss during a signal transmission through a fading medium. Measurements of this quantity have been made by experimenters using the hourly median value of the power available at the receiving antenna.

The fade duration is a random variable, and little is known about its statistical properties. The fading rate, which is derived as the number of times per unit time the received signal envelope crosses its median value with a positive slope, is used as a measure of the number of fades per unit time.

In addition to obtaining fading statistics, it is possible, and indeed more desirable, to directly measure error rates, error-free intervals, and distributions of the lengths of error clusters. Such statistics permit a more direct application of techniques for evaluation by analysis and simulation.

Unfortunately, there is a lack of unity in the various measurements reported in the literature. The same parameters were rarely obtained with the same ground rules or over the same intervals.

3.2.1 Land Lines and Submarine Cables

Data transmission via wire and cable is covered extensively in the literature. Frequency shift keying, coherent-phase shift keying, differentially-coherent-phase shift keying, and AM suppressed carrier are all widely used. These media are not subject to multipath fading, but are subject to impulse noise and dropouts which interrupt long error-free periods.

The tests reported in the literature operated mainly at 600, 1200, and 2400 bits per second with overall error rates ranging from 10^{-4} to 10^{-8} resulting from higher density clusters of errors.

NATO measurements on telephone lines at 5100 bits per second found overall error rates of 2.3×10^{-5} to 5.5×10^{-5} [28]. Tucker, Duffy, and Salva [29] conducted an extensive investigation and found error rates from 10^{-4} to 10^{-8} independent of the actual bit rates or modems. Lincoln Laboratory tests on the Hawaiian cable [30] at 1300 bits per second indicate an average bit error rate of 1.4×10^{-7}, with error-free periods of 72 hours.

Numerous other testing programs found similar results with typical error rates of 10^{-5} or better for ordinary telephone lines and 10^{-7} for cables [31-36]. These are overall error rates and are not descriptive of the burst on these channels. Indeed, the Kingston-Cape Kennedy tests showed that 37 percent of the error bursts exceeded one minute [34]. The average error rate in words containing errors was 31.75 percent. Daily variations in overall error rate of two orders of magnitude were common. In tests from Charlotte, North Carolina, to Miami, Florida, the average error rate in words containing errors was found to be 19.81 percent, with 31 percent of the bursts exceeding one minute [35]. Townsend and Watts found the error rates for long hauls to be four to five times higher than for short hauls [36].

3.2.2 Line-of-Sight Microwave

Performance data on line-of-sight microwave links is somewhat limited since most of the tests and measurements included cables or tropo links.

Coherent PSK and FSK are both used extensively on microwave links reported [29].

Fading caused by weather conditions bending the beam as well as by ground-reflected and airplane-reflected multipath has been observed [37, 38]. Space and frequency diversity have been helpful against the latter types of fading. Seasonal variations of 5 to 10 db in signal level have been observed.

Fading is severe during periods in which the atmosphere contains water vapor. There is also a frequency dependence, with deeper fades occuring at higher frequencies. According to experiments performed at 4000 mc on a 30.8 mile path, all the deep fades observed were frequency selective. During deep fades, 6-10 db losses that were not frequency selective were observed over a bandwidth of several hundred mc [39].

Although fades in excess of 30 db have been observed, the signal level dropped more than 7 db in summer only 1 percent of the time. In the winter the signal level dropped 5 db or more only 0.1 percent of the time [40]. Overall error rates of 5×10^{-5} or less with an error rate of 0.36 or less within erroneous blocks are typical [41].

3.2.3 Tropospheric Scatter

The theoretical aspects of tropospheric scatter transmission are covered extensively in Sunde [42]. FSK, PSK, and double sideband AM are indicated. Several extensive testing programs have been conducted.

Bullington, Inkster, and Durkee [18] report the results of their measurements over a 150-nautical-mile path using 505 mc and 4090 mc and over a

255-nautical-mile path using 505 mc. Measurements were obtained over irregularly spaced runs throughout a two-month period, with the duration of each run, T_{int}, taken to be 15 minutes. To eliminate the variations in median signal levels from run to run, the signal level was expressed in db with respect to the median level of the signal during the particular run. They found that most of the runs resulted in an approximately Rayleigh distribution. As examples of their findings, fades at 505 mc, 15 db below the median signal level, and more than one second in duration occurred approximately 35 times per hour; 10 db fades lasting more than one second occurred approximately 70 times per hour. It is interesting to note that at 4090 mc the fast fades occured about eight times faster. Hourly median signal levels over a period of one year at 505 mc and 4090 mc on the 150-nautical-mile path were also measured. The measured hourly median signal levels were converted into the path losses in excess of the free space loss and expressed in terms of db. This is a reasonable way to normalize the hourly median signal levels, but one has to exercise care in choosing the antenna gains to compute the free space loss. As expected, the results are reasonably close to log-normal with standard deviations of 9 db and 7.5 db for the 505 mc and 4090 mc cases, respectively.

The variation in the median signal level is due mainly to seasonal variations. Less loss is incurred during the summer and fall in comparison to winter and spring. There is also evidence of diurnal variations, but these appear to be less significant than seasonal variations for the experiment under discussion.

For a 255-nautical-mile path, the hourly fading depth was measured for a carrier at 505 mc [18]. The measurement was made by transmitting 500 watts at 505 mc. A radio receiver incorporating a 20 kc narrowband IF a- fier was used. The receivers measured the magnitude of the signal power de- livered by the receiving antenna. These measurements are summarized below.

Hourly Fading Depth (winter measurement) x db	Percent of Time Fading Depth is Less Than x db
7	1
10	10
11	20
15	65
18	90
22	99.5

For the same path and frequency, the number of fades per hour was related to the duration of the fades, as shown below.

Magnitude of Fade Below Hourly Median Value (db)	Fade Duration (sec)	Number of Fades per Hour with Longer Duration
15	1.0	32
	1.2	20
	1.5	10
	1.8	5
	2.0	3
10	2.0	30
	2.5	20
	4.0	10
	6.0	5
5	3.0	32
	4.8	20
	8.0	10
	16.0	5
	35.0	2

Other experiments performed in the VHF and UHF operating frequency ranges show a strong dependence of the fade rate on the frequency [43]. Using 5° beam antennas, the following measurements were obtained:

Carrier Frequency	x(Fades per Second)	Percent of Time the Average Fades Per Second Exceeds x
2290 mc	1.42	11
	1.20	15
	1.02	38
	0.9	50
	0.4	70
	0.3	80
	0.15	90
	0.10	95
417 mc	0.35	5
	0.20	12
	0.18	21
	0.13	40
	0.08	62

It was found that for UHF, a fade 6 db below the median value has an average duration of one second and at SHF an average duration of 0.1 second [43].

In the case of tropospheric scatter an aircraft can cause rapid and deep fades. These fades have been observed at about 30 cps at UHF to hundreds of cps at SHF, depending upon the geometry. To make matters worse, when the signal reflected from the aircraft is approximately equal in level to the tropospheric scatter signal, deep fades result.

The slow fade statistics obtained in [18] are indicative of gradual changes in the tropospheric scatter channel. The hourly median signal level was measured and its cumulative distribution obtained for both the 505 mc and 4090 mc channels. The hourly median path loss is then converted into the hourly median path loss in excess of the expected free space loss. The latter quantity is expressed in terms of db, and its cumulative distribution is shown in the table below for the 505 mc channel.

Hourly Median Path Loss in Excess of Free Space Loss x db	Percent of Hours that Hourly Median Path Loss in Excess of Free Space Loss Exceeds x db
86	2
78	10
72	30
66	50
62	70
55	90
47	98

This data, shown in very condensed form here, can be represented by a log-normal distribution.

The existence of seasonal variations can be seen from the chart below.

Month	Monthly Median of Hourly Median Loss Referred to Free Space Loss (db)
November	67
December	73
January	76
February	72
March	77.5
April	75
May	66
June	58
July	60
August	61
September	59
October	64

Systematic diurnal variations were noticed but were not significant. In terms of the hourly median values, there was an improvement of several db during the midnight to noon hours compared with the noon to midnight hours. This slight diurnal phenomenon was more pronounced in the summer than in the winter.

Several other investigators agree that the rapid fading was Rayleigh with 1/10 to 10 fades per second, and the slow fading was log-normal with seasonal variations [38, 44].

The White Alice system in Alaska is the largest tropo system in operation. Extensive tests found the rapid fading to be Rayleigh and the slow fading to be log-normal, with a standard deviation of 6 db [45]. On this system, quadruple diversity provided a 7 db advantage over dual diversity which, in turn, had a 14 db advantage over no diversity [46]. Overall error rates ranged from 10^{-2} to 10^{-6}. These were clustered, as indicated by the fact that the average number of errors in an erroneous 16-bit block was 5.26. In addition, the distribution of consecutive minutes in error indicates that 20 percent of the fades exceeded one minute and 2 percent exceeded four minutes. Also, 8 percent of the occurrences of consecutive error-free minutes exceeded 100 minutes [47].

A USAEL report on UNICOM tests found that for 50 percent of the time the error rate was less than 10^{-7}, and that for 50 percent of the time the intervals between errors were 27 minutes or longer [31].

3.2.4 Knife-Edge Diffraction

Statistical descriptions of knife-edge diffraction links are scarce. In most cases, they are mixed in with tropospheric scatter links. Indeed, about half the beyond-the-horizon paths in the White Alice system (see subsection 3.2.3) used knife-edge diffraction. It was noted that these links exhibited little of the deep fading phenomenon [46].

An Armour study revealed that for knife-edge diffraction the received signal is almost free of fading except in the presence of a ground-reflected signal [39]. With negligible ground reflection, a fade range of ± 3 db was observed over a 24-hour period. With ground reflection, a fading range as high as 10 db was observed.

3.2.5 Ionospheric Scatter

Ionospheric scatter communications are subject to severe rapid fading requiring space diversity to make channels usable. Frequency shift keying has gained virtually universal acceptance [43]. Tests made by the SHAPE Air Defense Technical Centre found error rates ranging from 2×10^{-3} to 4×10^{-5}. The rapidity with which the signal level fell from usable to unusable was remarkable. The satisfactory performance of detection-retransmission systems was also noted [31].

The hourly system loss has been measured in the frequency band between 25 mc and 108 mc (VHF) over a path length of 1000 to 2000 km and found to extend between 140 and 210 db. [43]. The power available at the receiving antenna was obtained from the hourly median value of a fluctuating received signal power. This provides a guideline to the extent of fading over a range of frequencies and path lengths.

The measured values of the transmission loss indicate that the extent of fading is frequency dependent. The transmission loss is the ratio of the power radiated from the transmitting antenna to the power available from the receiving antenna, and is proportional to f^n. The value of n depends on the transmission path and the time at which the measurements are obtained. Over the frequency range 30-108 mc, and for a particular transmission path, n was found to vary between 6 and 10. For system performance calculations, a value of n = 7.5 was found to be useful [43].

The short-time statistics of the fluctuating signal received over the same path reflect the frequency dependence observed in the transmission loss measurement. In measuring the received signal level, the receiving antenna beamwidth was chosen to have a gain of 20 db with respect to the isotropic in order to reduce the effects of meteor scatter. For various frequencies that fall within the frequency range of interest, the probability that the various carrier envelope levels in dbw are exceeded during the daytime is given below:

Frequency	Received Carrier Envelope Level x dbw	Probability that x dbw is Exceeded
30 mc	-135 -144	0.90 0.99
40.9 mc	-133 -137 -146 -157	0.10 0.50 0.90 0.99

Frequency	Received Carrier Envelope Level x dbw	Probability that x dbw is Exceeded
73.8	-149 -154 -161 -168 -179	0.01 0.10 0.50 0.90 0.99
107.8	-164 -168 -174 -182	0.01 0.10 0.50 0.90

From these statistics the fading range (power) can be obtained:

Frequency (mc)	Fading Range (power) (db)
40.9	13
73.8	14
107.8	14

If the fading range (power) for a random variable with Rayleigh distribution is obtained, the result is approximately 14 db. The fast fading characteristic during the day in this case is approximated closely by the Rayleigh distribution.

The fading rate of the envelope depends upon the signaling frequency. The following are typical:

Carrier Frequency	Fading Rate
30 mc	0.3 to 1.0 cps
50 mc	0.2 to 5.0 cps
108 mc	1.0 to 2.0 cps

The fading rate increases when antennas with wider beam widths are used, for the effects of increased meteor scatter are then included. Also, there is increased fading when transmission paths off the great circle path are used.

Other investigators found Rayleigh fading at 0.2 to 3 cps and log-normal slow fading with a standard deviation of 6 to 8 db for the hourly median [38,44].

Measurements performed to determine the diurnal variations encountered in typical ionospheric scatter links are shown below [43]. Data was obtained over a 1243 km path between Cedar Rapids, Iowa, and Sterling, Illin's, using a carrier frequency of 29.80 mc in December 1951. The upper and lower decile values refer to values in terms of system loss exceeded 10 percent of the days and 90 percent of the days, respectively.

Hour	System Loss Upper Decile (db)	System Loss Median (db)	System Loss Lower Decile (db)
00	194	189	184
02	192	187	180
04	189	187	182
06	190	190	185
08	188	185	180
10	186	178	171
12	180	174	168
14	184	175	167
16	189	182	173
18	193	189	181
20	195	190	185
22	194	191	186
24	193	189	183

As shown above, system loss is minimum during mid-day. In terms of the median value of system loss, there is a 15-db spread over a 24-hour period. There is maximum system loss in terms of the median value at about 2000 to 2200 hours at path mid-point.

Similar measurements were made in March and June over the same path, and it was found that minimum diurnal system losses occurred around mid-day both in the summer and winter. However, the median diurnal system loss around mid-day was about 10 db less in June than in March.

3.2.6 Ionospheric Reflection

HF radio channels using FSK, PSK, and SSB-AM are widely reported. However, detailed statistical data on fades or error distributions is scarce.

A report by Goldberg [48] describes tests over a 5000-mile path in which the transmitter power was varied. Median bit error rates ranged from 2.7×10^{-3} to 2.4×10^{-6} for FSK and 7.6×10^{-5} to 5.2×10^{-5} for PSK. Error bursts, but not solid clusters, were observed. Unicom tests found long periods with error rates above 10^{-2} [31].

Using dual diversity, multipath fading resulted in error rates of 10^{-2} to 10^{-3} [41]. The probability of an error-free one-second interval over a variety of 200-mile or shorter paths was found to range between 0.000 and 0.998 at 75 bps and 0.00 to 0.45 at 600 bps [37].

Very few measured fading parameter values appear to be available. Fading depths of 20 db have been observed in the operation of the AN/FGC-5 [49]. Another series of tests over 50-200-mile links found fade depths of 6 to 10 db, with a median fade duration of 8 msec—10 percent exceeding 17 msec and 1 percent exceeding 60 msec [50]. Other investigators found that the rate increased with frequency from 1.25 to 15 cps, with the average being 4 to 5 cps [44,51].

There is very little specific information concerning slow skywave fading other than some measurements pertaining to the day-to-day variation of the absorption fading phenomenon. It has been found that the day-to-day variations of the ionospheric absorption produce a lower decile value 6 db below the monthly median field intensity and an upper decile value 6 db above the monthly median.

3.2.7 Satellites

The current status of satellite channels is covered quite completely by Biterbi [52] and Froehlich et al [25]. Phase shift keying is universally accepted for satellite channels with either coherent or differentially coherent detection. Since the technology has been advancing so rapidly, only a limited amount of data is available on error statistics of satellite links whose properties are similar to those currently in use.

Froehlich et al [25] found one error in 10^9 bits over Telstar in a very short test. Bailey [26], working with Syncom, found transmission at 2400 bps essentially error free and at 50,000 bps found predominantly random errors at rates from 10^{-7} to 10^{-5}. Most of the errors were attributed to shortcomings of the terminal equipment rather than to problems with the satellite or disturbances in space. Although the error rate went up during thundershowers, the errors did not occur in long high-density bursts.

3.3 RELATIONSHIPS BETWEEN CHANNEL PARAMETERS AND SYSTEM PARAMETERS

An effective adaptive communications system must be carefully matched to the parameters of the channel in which it is to operate. Considerations of the maximum differential propagation delay, bandwidth of the medium, doppler frequency shifts, and phase variations lead to allowable values for the duration of a transmitted bit and the size of the alphabet. The nature of the errors to be encountered determines the choice of error control parameters. Once system parameters are matched to channel parameters, it remains to determine the thresholds for switching among the various sets of parameters.

3.3.1 Maximum Differential Mu' path Propagation Delay, Tm

3.3.1.1 Ionospheric Reflection

The path lengths through the ionosphere are continuously changing due to fluctuations in the altitudes of the ionospheric layers and movements associated with irregularities in the ionosphere. Under these conditions, the path length delay can be as high as several milliseconds [39] with delays up to 12 msec reported by the National Bureau of Standards in the 4 to 30 mc range [49]. In most cases T_m is between one and four milliseconds.

3.3.1.2 Ionospheric Scatter

Three types of multipath phenomena are observed in ionospheric scatter propagation accompanied by different delay characteristics [43]. Meteoric and auroral multipath echoes and ground-scatter waves propagated by the F-2 layer can be observed. The last phenomenon gives rise to delays of the order of 12 to 60 msec. The delay incurred from meteoric multipath is of the order of 1.0 msec, and the corresponding delay due to auroral multipath can be as high as 4.0 msec.

3.3.1.3 Tropospheric Scatter

The difference in arrival times of signals originating from the transmitter antenna's 3 db and 0 db points, and received through the receiver antenna's 3 db and 0 db points, respectively, can be computed from the separation of the transmitter and receiver and the gain of the antennas, and approximates the maximum differential delay. Bullington [53] does this and finds that for a 30 mile path with 40 db antennas, the maximum computed differential delay is 16 nanoseconds. Measurements were made under these conditions at 4,000 mc and the maximum differential delay was found to be 6 to 10 nanoseconds.

It has been predicted that, as the range is increased to beyond the horizon transmission paths, this delay will increase more rapidly. Tests have revealed that the maximum delay of the strong echoes could not be resolved by the equipment used, which could resolve 100 nanosecond delays. These tests were performed at ranges of 75-200 miles using antennas with 35 - 40 db gains and a pulsed signal frequency of 3700 mc.

3.3.1.4 Knife-Edge Diffraction

Little information has been found for this case, but the maximum delay due to multipath is expected to be small. This is based on a report [39] concerning fading measurements over a 160 mile path with an 8800 foot obstacle. Using a 38 mc signal, the received signal level varied from its mean by less than ±2 db over a 30-day test period.

3.3.2 Medium Bandwidth

Most of the media being considered exhibit frequency selective fading caused by multipath transmission. This phenomenon imposes a bandwidth limitation on the transmitted signal to avoid distortion of the received signal.

The following oversimplification should give an intuitive feeling for the medium bandwidth. Suppose a sinusoid with frequency f_0 cps propagates through the medium over two path lengths d_1 and d_2 and that the two signals are received having approximately equal intensity and a phase difference which is a multiple of 2π. Then

$$f_0 = \frac{m}{(d_1 - d_2)/c} \qquad (3\text{-}17)$$

where c is velocity of propagation and m is an integer. In this case, the received signal level is twice that of single path case. Next, suppose f_1 is a frequency for which the phase difference of the received signals propagated through the two paths is $m2\pi - \pi$ radians. Then

$$f_1 = \frac{2m - 1}{2(d_1 - d_2)/c} \qquad (3\text{-}18)$$

In this case the received signals are out of phase by π radians and cancel. Similarly, assume that f_2 is a frequency for which the received signals propagated through the two paths differ in phase by $m2\pi - \pi$, Then

$$f_2 = \frac{2m + 1}{2(d_1 - d_2)/c} \tag{3-19}$$

Thus, we have

$$f_2 - f_1 = \frac{1}{(d_1 - d_2)/c} \tag{3-20}$$

Some workers in the field use the above expression to define the bandwidth of the medium. However, this is not uniformly accepted because the assumption that only two paths exist is an oversimplification. In general, many paths exist and change with time, resulting to a medium bandwidth which is a random variable. A similar definition of the medium bandwidth is based on the requirement that the phase difference of the signals in the band propagated through two paths will be one radian at most. Hence,

$$2\pi f \left(\frac{d_1}{c}\right) - 2\pi f \left(\frac{d_2}{c}\right) = 1 \text{ radian} \tag{3-21}$$

Therefore

$$f = \frac{1}{2\pi\left(\frac{d_1 - d_2}{c}\right)} . \tag{3-22}$$

An average bandwidth is defined by other workers by measuring the cross-correlation coefficient of the received signal components separated in frequency. The difference between the frequencies that gives rise to a crosscorrelation coefficient of 0.5 is taken as the bandwidth of the medium.

3.3.2.1 Ionospheric Reflection

Diurnal variations of the optimum traffic frequency (by international convention, FOT) require that the frequency of transmission be changed several times a day to operate near the FOT. Different multipath delays are incurred with each frequency change as shown in the following table for a North Atlantic AM communication system with an F_2 layer maximum usable frequency (MUF) of 35 mc.

Frequency	Time Delay (ms)	Bandwidth (kc)
0.85 MUF = FOT	0.25	4
0.50 MUF = FOT	1.0	1
0.42 MUF = FOT	3.0	0.3
0.34 MUF = FOT	6.2	0.16
0.28 MUF = FOT	8.3	0.11

When operating at a frequency of 0.85 MUF, the available bandwidth is limited to 4 kc by selective fading.

The usable bandwidth can be increased by using space and/or frequency diversity techniques to eliminate the effects of selective fading. Frequency diversity techniques have been utilized to obtain total bandwidths of 16 or 20 kc, thus achieving information bandwidths of 8 to 10 kc. With frequency diversity, 32 teletype channels, each 250 cps wide, can be obtained. By adding space diversity the information bandwidth can be doubled.

3.3.2.2 Ionospheric Scatter

The medium bandwidth has been estimated to be 50 to 100 kc if the effects of meteoric, auroral and back scatter echoes can be reduced [51]. This estimate is based on computed delay times due to multipath transmission. Under certain conditions this has been estimated to be nine microseconds, but 20 microseconds has been considered representative. In practice, a bandwidth of this magnitude has not been achieved. It has been found that the cross-correlation of received signals separated by 5 to 10 kc is 0.5. Using frequencies sufficiently high to minimize the delayed echoes from backscatter, the medium has a bandwidth of 2 to 4 kc greater.

Measurements made by others support the information above [43]. A 6° narrow beam antenna was used to make the following measurements at 50 mc.

Time	Correlation Coefficient	Received Signal Frequency Separation
Early Morning	0.5	6 kc
Midday	0.65	6 kc
Midday	0.5	7 kc

When a broader beamwidth antenna is used, or off-great-circle transmission paths are used, the crosscorrelation coefficient is smaller.

3.3.2.3 Tropospheric Scatter

Analytical expressions have been obtained for the medium bandwidth, depending upon whether the receiving antenna beamwidth is broad or narrow. When the antenna 3 db point, expressed in radians, is greater than $\frac{2}{3}\frac{D}{a}$, where D is the transmitter to receiver distance and a is the effective earth radius, then the antenna is broad beam. The properties of the troposphere are such that a horizontal beam width greater than $\frac{2}{3}\frac{D}{3}$ does not add to the effectiveness of the system. For the broad beam width case, the medium bandwidth in mc is

$$B = \frac{4}{\left(\frac{D}{100}\right)^3} \tag{3-23}$$

Experimental results indicate that for a path length of about 200 miles and an antenna of 1.2 degrees in beamwidth, the medium has a bandwidth of approximately 10 mc. The expression given for the broad beam case has also been found to agree with experimental results.

The tropospheric bandwidth was also determined for a different path using crosscorrelation coefficient measurements [43]. The transmitted signal frequency was 2290 mc, the path length 188 miles, and the antenna beamwidth one degree. The transmitted signal was swept over a 16 mc band at 2290 mc, and the crosscorrelation coefficients of the signals received at 2290 mc and at offset frequencies were measured. It was found that, at an offset of 4 mc from 2290 mc, there was practically no correlation.

3.3.2.4 Line-of-Sight Propagation

Multipath transmission determines the bandwidth in this case. A differential path delay corresponding to seven feet was obtained for a 22-mile path at 4000 mc using transmitting and receiving antennas of 34-to 40-db gain. Taking the reciprocal of the delay gives a bandwidth of 140 mc. In practical systems, the bandwidths are limited to 20 to 30 mc.

3.3.3 Other Channel Characteristics

The two other channel parameters to be considered here, doppler frequency shift and phase variation, are particularly relevant to the case of ionospheric scatter propagation. However, the discussion is not limited to the ionospheric scatter case.

The transmission of a sinusoidal signal results in a doppler frequency offset of the received signal in ionospheric scatter propagation. This is primarily due to the presence of ionization trails, and according to doppler frequency shift measurements made on a VHF path at 50 mc, doppler frequency shifts as high as 4700 cps have been observed [43].

The doppler frequency spread bandwidth is taken to be two times the maximum doppler frequency shift expected under the preceding conditions.

The integrated phase variation has been examined using a basic integration time of 20 msec. The variation is obtained by taking the difference in integrated phase from one 20 msec interval to the immediate preceding one. An ionospheric scatter propagation path from Long Branch, Illinois, to Boulder, Colorado, using frequencies of 30 and 49.6 mc, respectively, exhibited the following distributions of phase variations:

Phase Variation ($\phi°$)		Percent of Time Phase Variation is Greater than $\phi°$
30 mc	5	16
	15	4.4
	25	2.4
	35	1.2
	45	0.9
	55	0.67
	65	0.5
	84	0.38
	105	0.27
	114	0.23
	144	0.1
49.6 mc		
	5	60
	15	26
	30	8.5
	45	3.5
	65	1.8
	85	0.92
	105	0.66
	125	0.32
	145	0.20
	175	0.1

The perturbation associated with the received phase angle is not confined merely to ionospheric scatter propagation. An HF ionospheric reflection measurement over a transcontinental path was performed by transmitting a continuous sinusoid at 15 mc and making a simultaneous recording of the amplitude and frequency of the received signal [18]. The relative phase was obtained from the doppler frequency recordings by an integration process. Examining the phase variation over periods of 20 msec, the maximum phase variation is of the order of 10 degrees.

Further experimental measurements concerning phase fluctuations are reported by Brennan and Phillips [54], Hagfors and Landmark [55] and Voelcker [56].

3.3.4 Constraints on Bit Duration

3.3.4.1 PSK

When the bit duration is T, the PSK communication system bandwidth is proportional to T^{-1}. Thus, a system that has a bandwidth narrow in comparison to the coherent bandwidth of the medium will satisfy $T > T_m$. This is a necessary condition for reducing the effects of interfering signals received through multipath propagation during a bit interval.

When the system bandwidth is narrow in comparison to the coherent bandwidth of the medium, a carrier transmitted at the center of the channel will be received as a sinusoid whose amplitude and phase fluctuate slowly.

The phase stability parameter can be generally discussed in terms of the doppler frequency bandwidth B_s and the information bit duration T. This can be seen by considering the maximum phase variation in radians due to the maximum doppler frequency shift, $B_s/2$ cps over a T second interval. This maximum phase variation is obtained by integrating πB_s over the duration T to get $\pi B_s T$. If $B_s T < 1$, the phase variation over T seconds can be made relatively small.

The constraint required to maintain a small phase variation over the bit duration is thus $T < \frac{1}{B_s}$. Combining the two requirements,

$$T_m < T < \frac{1}{B_s} . \tag{3-24}$$

The phase stability of the medium as well as the maximum path length delay become critical when phase shift keying is used. To reduce the phase variation due to the medium, the bit duration can be reduced, but the limiting factor is the maximum path length delay. When the multipath delay becomes a significant portion of the bit duration, the interference due to multipath can be intolerable and result in system performance deterioration.

In considering the use of M-ary phase angles for modulation, similar considerations are involved as for the case of biphase modulation. There is a maximum phase angle variation due to the transmission medium that can exist and still have the system meet the requirements on the bit error rate. When the maximum phase variation due to the medium is considered, the effects of receiver noise must also be incorporated. When this is done, and the required error rate specified, meaningful bounds can be imposed upon the maximum phase variation for certain slow Rayleigh fading.

3.3.4.2 FSK

In the case of FSK, care must be exercised to maintain a sufficient frequency separation between the mark and space spectrum of the order of $\frac{B_s}{2}$. This value will depend on the medium and the frequency at which the system operates.

The bit duration T is generally chosen to satisfy the previously mentioned constraint that $T > T_m$, but many variations have been devised to minimize the interference due to multipath delay.

T can be chosen greater than $\frac{1}{B_s}$. When this is done T should include at least several subintervals of the order of $\frac{1}{B}$. In each subinterval, $\frac{1}{B_s}$, a small phase variation may occur, but over the total duration T the phase angle will tend to be random. For the detection of received signals extending over a duration $T > \frac{1}{B_s}$, the received signal energy can be measured at $\frac{1}{B_s}$ intervals and combined.

Interference due to multipath signals and the methods devised to improve digital communication system performance give rise to various FSK systems. Based on the expected multipath delay, a dead time is often provided after the mark or space symbols to prevent multipath interference. A system has been designed in which the AN/FGC-5 has been modified by providing three pairs of mark-space channels to operate in the presence of 12-msec delayed signals 20 db greater in amplitude than the signal received via the direct path [49, 57]. The use of the dead time in FSK systems and the use of additional mark-space subchannels to reduce interference due to multipath make it mandatory that frequency division and time division multiplexing of subchannels be performed in order to transmit at an efficient bit rate.

3.3.5 Parameters Related to Selection of Redundancy Technique

The redundancy methods under consideration here are general and include error control techniques as a special case. Diversity techniques, provision of a margin in the transmitter power requirements, existence of dead time following a mark or space symbol in FSK systems, variation in the duration of a bit or basic signaling element, use of an RF signaling bandwidth larger than that required for the information transmission rate, etc., are considered to be redundancy techniques exclusive of error control.

For the design of diversity techniques, the various correlation coefficients associated with the particular diversity methods are required. For example, in the case of space diversity the crosscorrelation coefficient of signals received through two antennas must be determined as a function of antenna separation in wavelengths for the particular space diversity technique. The separation that reduces the crosscorrelation coefficient to $\frac{1}{e}$ is the desired antenna separation.

The parameters associated with the depth of fade combined with the average error rate over a fade become pertinent in any discussion of transmitter power margin. In PSK systems the probability density function of the phase variation is independent of the power received. It is expected, then, that increasing the transmitter power would have limited influence in improving PSK system performance–this turns out to be the case.

The criteria for selecting the basic signaling interval T discussed above provide information on the constraints on varying the bit duration to provide more average energy per bit.

For PSK systems that employ coherent detection, Cahn [58] shows (Figure 3-1) the signal-to-noise ratio, measured in equivalent bandwidth, required to maintain a specified error rate when M-ary phase is used. This relation is

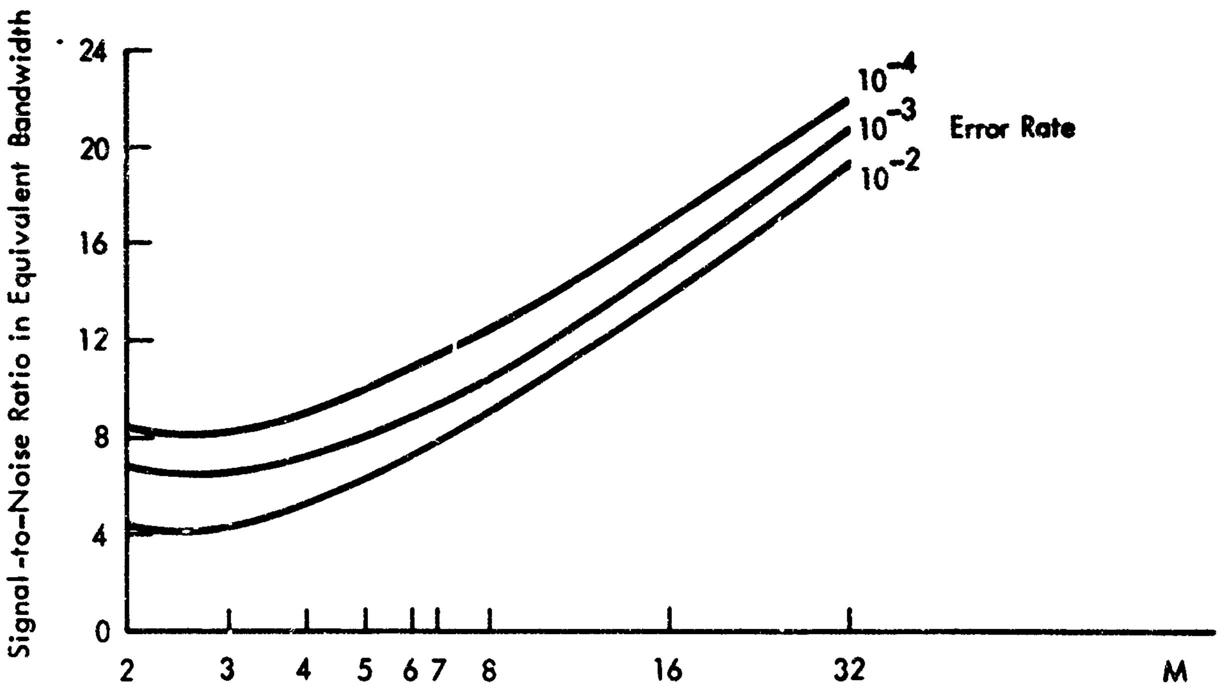


Figure 3-1. Signal-to-Noise Ratio in Equivalent Bandwidth versus M-ary Phase.

derived for the case where there is gaussian white noise and no fading. It is assumed in Cahn's result that the information rate is fixed, and the equivalent bandwidth is the bandwidth required in a biphase system to transmit at the specified rate. The performance degradation with differentially coherent detection is less than 3 db in the range of interest.

3.3.6 Selection of Coding Parameters

If distributions of burst lengths and error-free intervals are available, it is a straightforward matter to select parameters for a code to correct any desired fraction (less than unity) of the errors. Unfortunately, such input data is rarely available.

A procedure for selecting error control techniques for real systems from raw serial error statistics was developed by Benice and Frey [59]. Several shortcomings have been discovered in the procedures originally developed, but efforts to improve the original procedures have been successful.

A new estimate, ϕ (b, n), has been derived for the probability that all the errors in an erroneous n-bit block will be contained in a single b-bit subblock (where b-bit subblocks are considered to be placed end-to-end, starting with the first bit in the n-bit block).

Let f(n) denote the probability that an arbitrary block of length n will contain an error. Let P(k and $\bar{b}$) denote the probability that in an arbitrary block of length k+b bits, the first k bits will be correct and the last b bits will contain an error. Let P($\bar{b}$ and k) denote the probability that the first b bits will contain an error and the last k bits will be correct. Let P(k | $\bar{b}$) denote the conditional probability that the first k bits in a block of length $k+b$ bits will be correct given that the last b bits contain an error.

The probability that the first k bits of an arbitrary block of length $n > k + b$ will be correct, the next b bits contain an error, and the last $n-k-b$ bits be correct is given by

$$P(k \mid \bar{b})\, P(\bar{b} \text{ and } n-k-b), \tag{3-25}$$

provided that the conditional probability is independent of the fact that the last $n-k-b$ bits are correct. Although this assumption is not always valid, it does provide a close approximation to the correct probability and provides an expression which can be derived from measured block error rates as shown below.

Note that

$$f(k+b) = f(k) + P(k \text{ and } \bar{b}) \tag{3-26}$$

since an erroneous block of length $k+b$ bits must either have an error in the first k bits or have the first k bits correct and an error in the last b bits. Similarly we have

$$f(k+b) = P(\bar{b} \text{ and } k) + f(k) \tag{3-27}$$

It follows that

$$P(k \text{ and } \bar{b}) = f(k+b) - f(k) = P(\bar{b} \text{ and } k) \tag{3-28}$$

Thus we have

$$P(k \mid \bar{b}) = \frac{f(k+b) - f(k)}{f(b)} \tag{3-29}$$

Using (3-28) and (3-29) to obtain an alternate expression for (3-25), we have

$$P(k \mid \bar{b}) P(\bar{b} \text{ and } n-k-b) = \frac{f(k+b) - f(k)}{f(b)} (f(n-k) - f(n-k-b)) \tag{3-30}$$

Formula (3-30) applies to an arbitrary block of length $n \geq k+b$. To obtain the probability of the same condition for an erroneous block of length n, one need only divide the expression by f(n). This will be used to obtain the new estimate $\phi(b,n)$ for the probability that all the errors in an erroneous block of length n are contained in a single b-bit subblock. Consider first the case $n=ib$ where i is a positive integer. Now we have

$$\phi(b,n) = \sum_{j=1}^{i} \frac{P((j-1)b \mid \bar{b})\, P(\bar{b} \text{ and } (i-j)\, b)}{f(n)} \tag{3-31}$$

Using (3-30) this becomes

$$\phi(b,n) = \sum_{j=1}^{i} \frac{[f(jb) - f((j-1)b)]\,[f((i-j+1)b) - f((i-j)b)]}{f(b)\, f(n)} \tag{3-32}$$

Consider now the case $n=cb$ where $i \leq c \leq i+1$. Define $\phi(b,n)$ by

$$\phi(b,n) = \exp\left\{ (i+1-c)\, \ln \phi(b, ib) + (c-i)\, \ln \phi(b, (i+1)b) \right\} \tag{3-33}$$

Since the graph of $\phi(b,n)$ as a function of n is expected to be similar to a negative exponential, the interpolation given in (3-33) should be fairly smooth.

However, it is not the only possible interpolation that could be used, and no claim is made that it is the best one. This approximation $\phi(b, n)$ should provide appropriate inputs to the formula for $\theta(b, n)$ given in reference [59].

3.4 RELATIONSHIPS BETWEEN MODELS, FADES, AND ERRORS

The ultimate evaluation of adaptive data transmission techniques will be based, to a large degree, on the error rate in the data bits reaching the destination. It should be kept in mind that the channel models and fading statistics are merely devices used to approximate the necessary but unavailable error statistics or serial error stream.

The independent error channel is, of course, the simplest. By simply counting errors and total bits over such a channel the single defining parameter, the probability of error, can be determined. The model then will generate a channel having the same error statistics. Unfortunately, as we have seen, this model is representative of very few real channels, and is not a channel in which adaptive procedures can generally be of assistance (subsection 4.4).

The Gilbert model is next in complexity. It will generate directly a serial stream of error locations characterized by long error-free periods and high-density clusters of errors. Its three parameters can be found directly by measuring the distributions of burst lengths, error-free run lengths, and error rate within bursts. This model adequately represents land lines, which, however, are quite adequately served by detection-retransmission systems and thus do not need the more sophisticated techniques under investigation here. But it cannot represent the fading channel, which is the prime concern here. The fading channel is characterized by a fluctuating error probability within error bursts as well as a fluctuating longer term mean error probability.

The main interest, thus, is in the fading models–Rayleigh, Rician, or gaussian for rapid fading and log-normal for slow fading. The models do not immediately yield error patterns. They are merely distributions of the envelope of the received signal. From these the distribution of received signal power can be obtained immediately. But these are distributions and not Markov processes that give a serial description of the channel.

Rice [15] has derived some useful formulas and approximations which enable us to fill in part of the serial fading picture. If we can determine the rate $N(E_0)$ at which the envelope E crosses the level E_0 in a given direction by direct measurement or estimate it from the formula

$$N(E_0) = \sqrt{\frac{\beta_E}{2\pi}}\ \frac{E_0}{\Psi_0}\ \exp\left(\frac{-E_0^2}{2\Psi_0}\right) \tag{3-34}$$

where Ψ_0 is the mean square value of the narrow band noise and $\beta_E = 4\pi^2 \int_0^\infty w(f)\,(f-f_0)^2 df$, (where f_0 is the center frequency of the band-pass spectrum) is a second moment of the spectral density, then we can use the following formula for the ratio of the number of threshold crossings per second at level $E = E_0$ to the number of crossings at level $E = 1$:

$$\frac{N(E_0)}{N(1)} \approx 1.649\, E_0\, e^{-E_0^2/2} \tag{3-35}$$

This has the form of a Rayleigh density function and is plotted in Figure 3-2 along with the Rayleigh density function for the envelope

$$f(E) = \frac{2E}{\overline{E^2}} \exp\left(-E^2/\overline{E^2}\right) \tag{3-36}$$

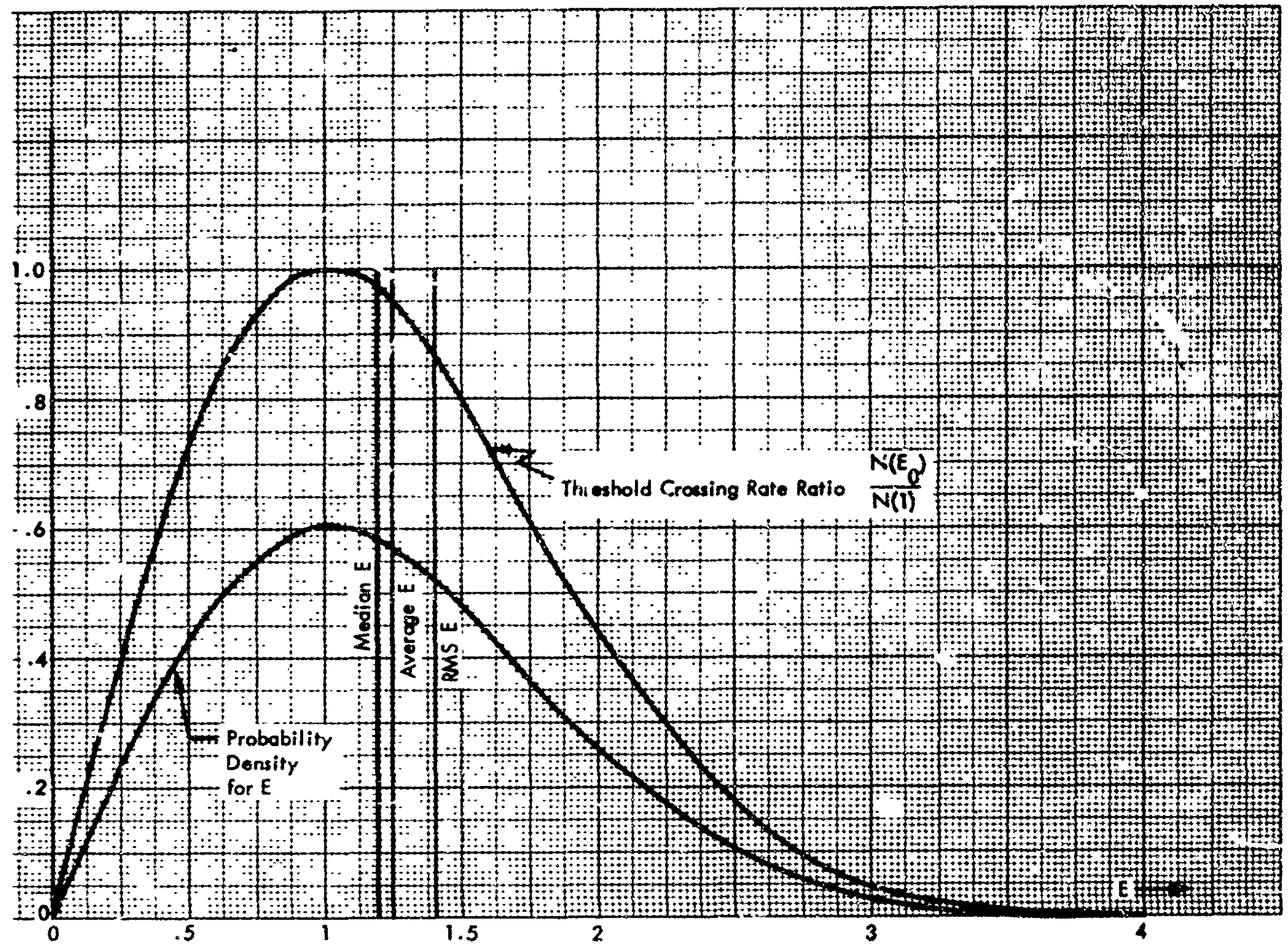


Figure 3-2. Threshold Crossing Rates (Envelope)

with $\overline{E^2} = 2$ (so that average power is unity). The same ratio is plotted as a function of received signal power $P = \frac{E^2}{2}$ in Figure 3-3 along with the exponential density function for received power.

To illustrate the use of these curves, suppose that at the mode of the envelope distribution corresponding to 0.5 milliwatt of received power the threshold crossing rate is 1000 per hour. Then at a threshold of 0.25 mw the threshold crossing rate would be approximately 900 per hour according to Figure 3-3, and at a threshold level equal to 3 mw the crossing rate would be approximately 190 crossings per hour (in either direction).

Looking ahead, it should be noted that, although the actual statistics for a real channel may be unknown, it is possible to determine the mean square value of the envelope of a Rayleigh distribution fitting the observations by finding the level for which the threshold crossing rate is a maximum. This gives the mode, M, whence the mean-square value of the envelope is $2M^2$. This will not necessarily be a best least-squares fit, but is reasonable in the sense that it fits a Rayleigh model to the threshold crossing rates, which are the most important aspects of the probability distributions of the threshold variable.

We can also compute the average duration of the intervals for which the variable θ (envelope or power) lies between θ_1, and θ_2 $(\theta_1 < \theta_2)$. Let N_1 and N_2 represent the number of times the thresholds θ_1 and θ_2, respectively, are crossed per second, and let ϕ_1 and ϕ_2 represent the fractions of time that the variable θ is greater than θ_1 and θ_2, respectively (as obtained from the probability distribution of θ). Then the average duration is given by

$$\overline{T_{12}} = \frac{\phi_1 - \phi_2}{N_1 + N_2} \qquad (3\text{-}37)$$

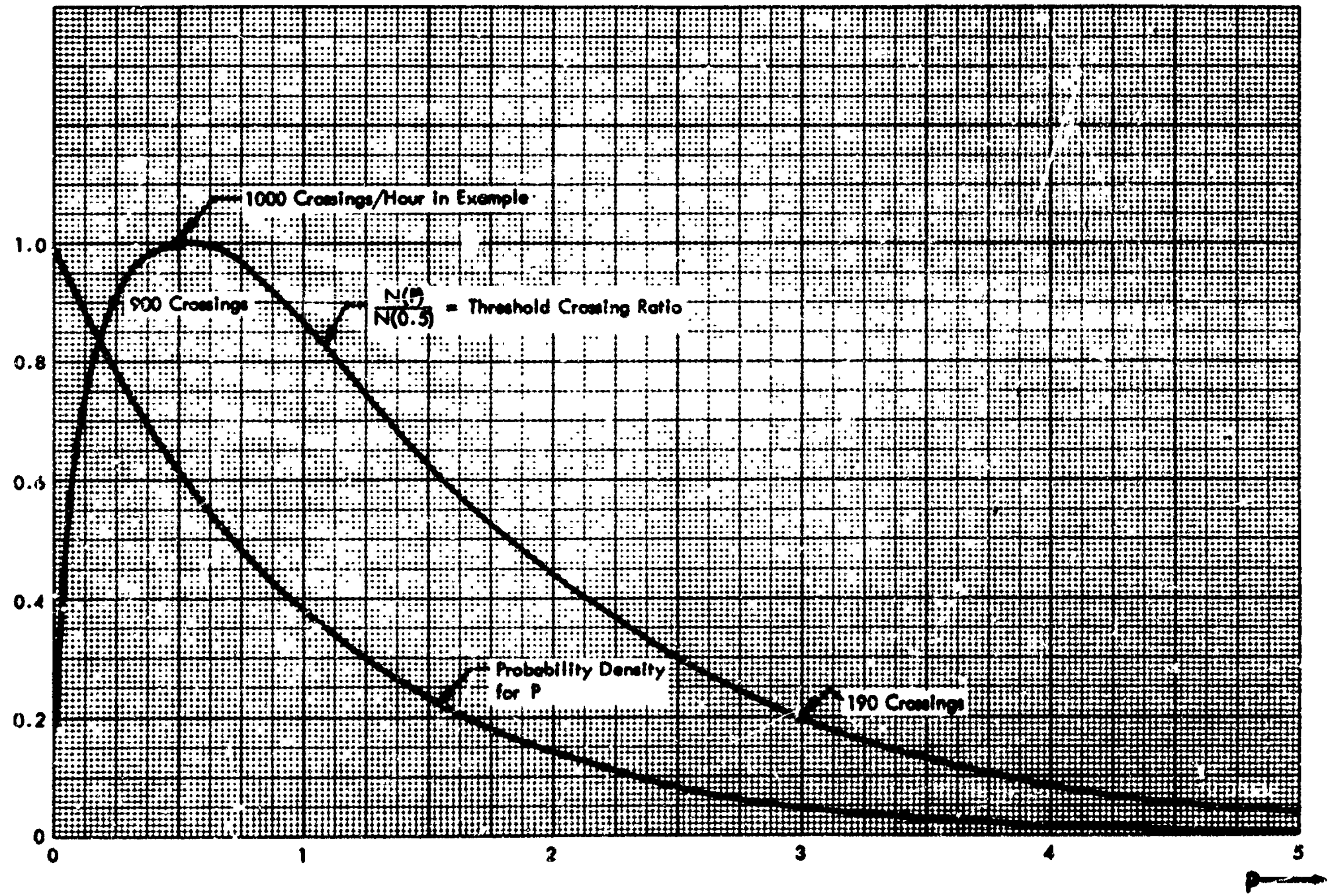


Figure 3-3. Threshold Crossing Rates (Power)

Formulas are readily available to give the instantaneous probability of error for a given modulation technique and signal to noise ratio. For example, with non-coherent FSK we have

$$P_e = \frac{1}{2} \exp(-E/2N_0); \tag{3-38}$$

for coherent PSK

$$P_e = \frac{1}{2}\left(1-\operatorname{erf} \sqrt{E(1-\rho)/2N_0}\right) \tag{3-39}$$

where ρ is the cross correlation coefficient of the signaling waveforms; and for differentially coherent PSK

$$P_e = \frac{1}{2} \exp(-E/N_0) \tag{3-40}$$

Formulas for longer term average error rates in channels subject to Rayleigh and other types of fading are available. These formulas were not used because this analysis required the identification of the exact nature and specific locations of the error patterns. Consequently, simulated fades had to be produced, and errors had to be located individually using the instantaneous error probability formulas (Equations 3-38 through 3-40).

For the channel models being considered, the defining parameters can be measured directly or can be computed directly from channel measurements. These aspects were considered in subsections 3.1.1 to 3.1.4, where the models were described and discussed. In instances where the channel measurements were inadequate for direct computations, we were able to find appropriate model parameters to fit the observations by trial and error computer simulations of the channels.

Section 4

THE DESIGN OF ADAPTIVE SYSTEMS

A multitude of problems must be considered in designing a practical adaptive data transmission system. How to evaluate the channel status in real time must be resolved. The nature of delays and their effects on performance must be ascertained. A control system must be developed that can operate in the presence of errors.

In addition, it must be decided whether a particular channel can be improved by the introduction of adaptive techniques. If a given channel is improvable, an optimum set of operating modes and an optimum set of times for changing modes must be found. The problems of bit, block, and mode synchronization plus buffering requirements must be resolved. Finally, the feasibity of building such systems at reasonable cost must be considered.

However, before these problems can be considered, a basic framework of terminology for describing and comparing the operation of adaptive systems as well as standards for comparing their relative performance must be developed.

4.1 BASIC PRINCIPLES OF DESIGN AND EVALUATION

In this section, the basic framework for describing, evaluating, and comparing adaptive data transmission systems is established. First, the concept of a threshold variable is introduced. This is a single, variable quantity from which the state of the channel can be determined. Next, a set of evaluation functions (which have the property that the performance of the system at any instant is directly proportional to their value) is introduced. The evaluation functions depend only on the mode of operation of the system and on the threshold variable. The statistical average of the appropriate evaluation functions for the several modes is shown to be equivalent to the time average, and this statistical average forms the basis of the overall performance index for a particular adaptive scheme. Refinements of the simple average of the performance index are shown to be necessary to account for inevitable delays and errors in making transitions from mode to mode.

4.1.1 Threshold Variables

The concept of an adaptive communication scheme implies that there is a changing state of the channel and that as the channel state changes, performance can be improved by changing the mode of operation of the system. The first problem faced by an adaptive scheme is thus the identification of the present "state" of the channel. Several quantities have been suggested as contributing factors in the definition of the system state, primarily received signal level, noise level, (also combined in signal-to-noise ratios), and observed error rates. The quantity most commonly referred to in connection with fading channels is the received signal-to-noise level (say, at the output of the IF section of a receiver), and this is really equivalent to the received signal power level, since the noise in most analyses is assumed constant (also usually assumed band-limited, white, and gaussian). For our purposes a single threshold variable will be assumed whose domain can be partitioned into intervals which correspond to discrete "states" of the system. The dividing lines between the intervals are called thresholds.

It will often be convenient in the sequel to consider the threshold variable to be simply the signal-to-noise ratio. However, any combination of measurable quantities which adequately serve to identify or recognize the system state is an allowable threshold variable. The most important idea here is that at each instant of time the state can be determined by computing a single number from measurable quantities.

4.1.2 Evaluation Functions

For every communication problem in which it is desired to adapt, there will be implicit certain constraints on the system. Within these constraints, however, there may be a wide choice of allowable schemes or sets of parameters. For each allowable scheme, an evaluation function will be formulated, which will take into account the probabilities of error, rapid fading effects, speed of transmission, and other measures of performance, all as a function of the threshold variable. The threshold variable is the single quantity, or statistic, upon which are based the adaptive decisions. The evaluation functions will typically be monotonic increasing functions of the received signal energy or signal-to-noise ratio (after diversity, if any).

The magnitude of the evaluation function will be directly proportional to the "quality" of the communication link at each instant, so that an overall performance index can be computed for an adaptive communications scheme by averaging over time the instantaneous values of the evaluation function. Thus the definition of "quality" is completely arbitrary, and is specified by the form of the evaluation function.

Several hypothetical evaluation functions (denoted by F's) corresponding to different modes of operation for a hypothetical system are illustrated in Figure 4-1. The threshold variable for this example is signal-to-noise ratio. For $S/N < \alpha$, Mode 2's, evaluation function is seen to have the largest value; for $\alpha \leq S/N \leq \beta$ Mode 3 is best, and for $S/N > \beta$, Mode 5 is best. An obvious first estimate would indicate that the system could be optimized by observing the threshold variable, S/N, and transmitting in Mode 2, Mode 3, or Mode 5 when the corresponding evaluation function is maximum; the thresholds on the S/N ratio would be α and β.

Typically, the class of all allowable schemes will have an upper bound to its members' evaluation functions, made up of segments from a great many evaluation curves. The upper bound curve is shown by a dotted line in Figure 4-1. Attaining the performance indicated by this upper bound may require numerous transitions from mode to mode as the signal-to-noise ratio fluctuates. It is then a problem of system design to determine how many transitions are optimum, and where the transitions should be made. The

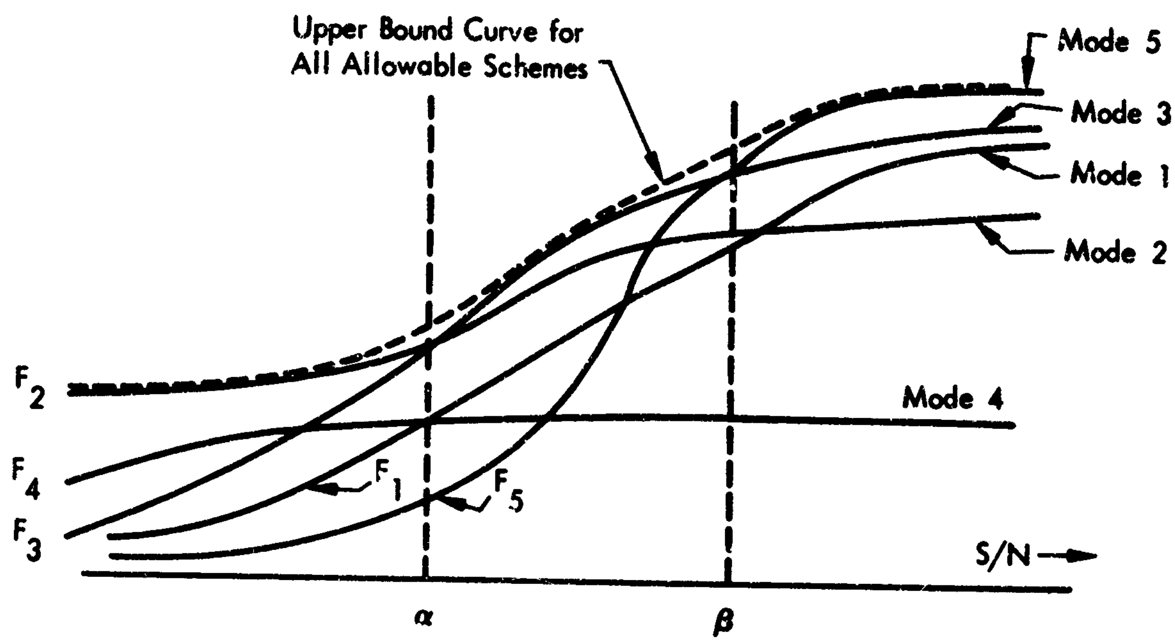


Figure 4-1. Hypothetical Evaluation Functions.

graphical display of the performance in various modes is an aid in deciding which modes should be utilized. For instance, in Figure 4-1 it may be unprofitable to use Mode 3, since its performance exceeds that of Modes 2 and 5 only slightly for values of S/N between α and β.

It is clear that a cost must be assigned to the act of making a transition, since otherwise it would be advantageous in all cases to follow the upper bound curve. Once two modes have been decided on, a logical placement of the threshold is at the intersection of the graphs of the evaluation functions. However, because of the non-zero cost of the transition, it may be desirable to delay making the transition until the threshold variable moves considerably past the intersection point.

4.1.3 Performance Indices

The evaluation function $F(\theta)$ forms a type of instantaneous performance index. When integrated over an interval of equally likely θ, $F(\theta)$ yields a performance index

$$I = \frac{1}{\theta_1 - \theta_0} \int_{\theta_0}^{\theta_1} F(\theta)\, d\theta = \int_{\theta_0}^{\theta_1} F(\theta) \cdot \frac{1}{\theta_1 - \theta_0}\, d\theta \qquad (4\text{-}1)$$

which is a figure of merit representative of the operation of the system over that interval. In general, the threshold variable will be described by its distribution density $f(\theta)$ and the basic performance index will be

$$I = \int_0^{\infty} F(\theta) f(\theta) d\theta \qquad (4\text{-}2)$$

For the situation illustrated in Figure 4-1, using three modes, the performance index is

$$I = \int_0^{\alpha} F_2(\theta) f(\theta) d\theta + \int_{\alpha}^{\beta} F_3(\theta) f(\theta) d\theta + \int_{\beta}^{\infty} F_5(\theta) f(\theta) d\theta \tag{4-3}$$

From this ideal value must be substracted the corrections needed because of inefficient operation in the wrong modes after each transition (see subsections 4.3 and 4.5). For this, it will be convenient to treat separately the intervals during which the system occupies a non-optimum state because we can treat statistically the "ideal" threshold crossings when transition should occur.

This correction will be incorporated in a fixed average cost per transition. The duration of occupancy in a new mode need not be considered in figuring the cost of a transition, for the cost is actually associated only with the actual making of a transition. Indeed, if there were no cost for transition, there would be no penalty for short duration in a mode. The durations do enter the system design, however, by affecting the choice of thresholds, since average durations which are too short will not have enough increased efficiency to offset the losses in efficiency due to transitions.

The revised performance index can then be found by considering the frequency of transitions, i.e.,

$$I = \int F(\theta)\, f(\theta)\, d\theta - \sum_{i=1}^{n-1} c_{i,\, i+1} N_i^{+} - \sum_{i=2}^{n} c_{i,\, i-1} N_i^{-} \tag{4-4}$$

where $c_{i, i+1}$ and $c_{i, i-1}$ are the costs per transition from the i^{th} to $(i+1)^{th}$ and $(i-1)^{th}$ modes respectively and N_i^+ and N_i^- are the average number of transitions per second from the i^{th} to $(i+1)^{th}$ and $(i-1)^{th}$ modes, respectively.

In a typical example, the integral above might yield an average information rate, in bits/sec; $c_{i, i+1}$ $[c_{i, i-1}]$ would be the equivalent loss in bits per transition to next mode (above +, or below -). The performance index above could also be interpreted as a dimensionless quality factor. In that case, the average number of transitions per unit time can be absorbed into an overall cost $C_i = N_i (c_{i, i+1} + c_{i, i-1})$ for each transition where N_i is the number of transitions per unit time in either direction if the thresholds for both directions are the same. Note that $N_i^+ = N_i^- = N_i$. Thus,

$$
\begin{aligned}
I &= \int F(\theta) f(\theta)\, d\theta - \sum_i N_i (c_{i, i+1} + c_{i, i-1}) \\
&= \int F(\theta) f(\theta)\, d\theta - \sum_i C_i
\end{aligned}
\qquad (4\text{-}5)
$$

where the index i is now associated with the thresholds between states and modes. In order to allow for the inevitable errors in transitions made by the system, the performance index could be generalized to incorporate the probabilities of mode occupancy by the system as a function of the threshold variable.

It may be practical to neglect the probabilities of "wrong-mode" occupancy, and to absorb the average loss in efficiency from such cases in the transition costs. Thus, every threshold crossing is considered to eventually lead to a transition, but with an average delay of some known duration during which the system's efficiency is impaired. The added loss in efficiency is then added to the other costs of the transition. Figure 4-2 shows the evaluation functions, upward and downward thresholds, and the distribution density of the threshold variable.

Figure 4-3 illustrates a hypothetical situation with three modes, showing S/N ratio and the time integral of evaluation and costs as a function of time. The transition costs in this representation appear as negative-going steps in the time-accumulated evaluation curve. The upward and downward thresholds are illustrated for each transition. Note that all transitions' costs are not equal, although all transitions of the same type (e.g., A to B, downward transition) are shown with the same average cost. In Figure 4-3, the performance index is the slope of the line from the origin to a point on the accumulated evaluation curve at some time in the distant future.

The two main performance indices are the pair (R, ρ) of throughput and error rate, and the information theoretic measure of mutual information (or transinformation) which in effect combines R and ρ into a single number. (It will be shown in subsection 4.1.4 that the information theoretic measure may not be suitable for evaluation purposes.)

The throughput, R bits per second, is a constant for any mode of operation which does not discard detected errors. The evaluation function F(θ) has the form:

$$R = F(\theta) = d(\text{digits per second}) \times \log_2 m(\text{bits per digit}) \times \left(\frac{n-r}{n}\right) \quad (4\text{-}6)$$

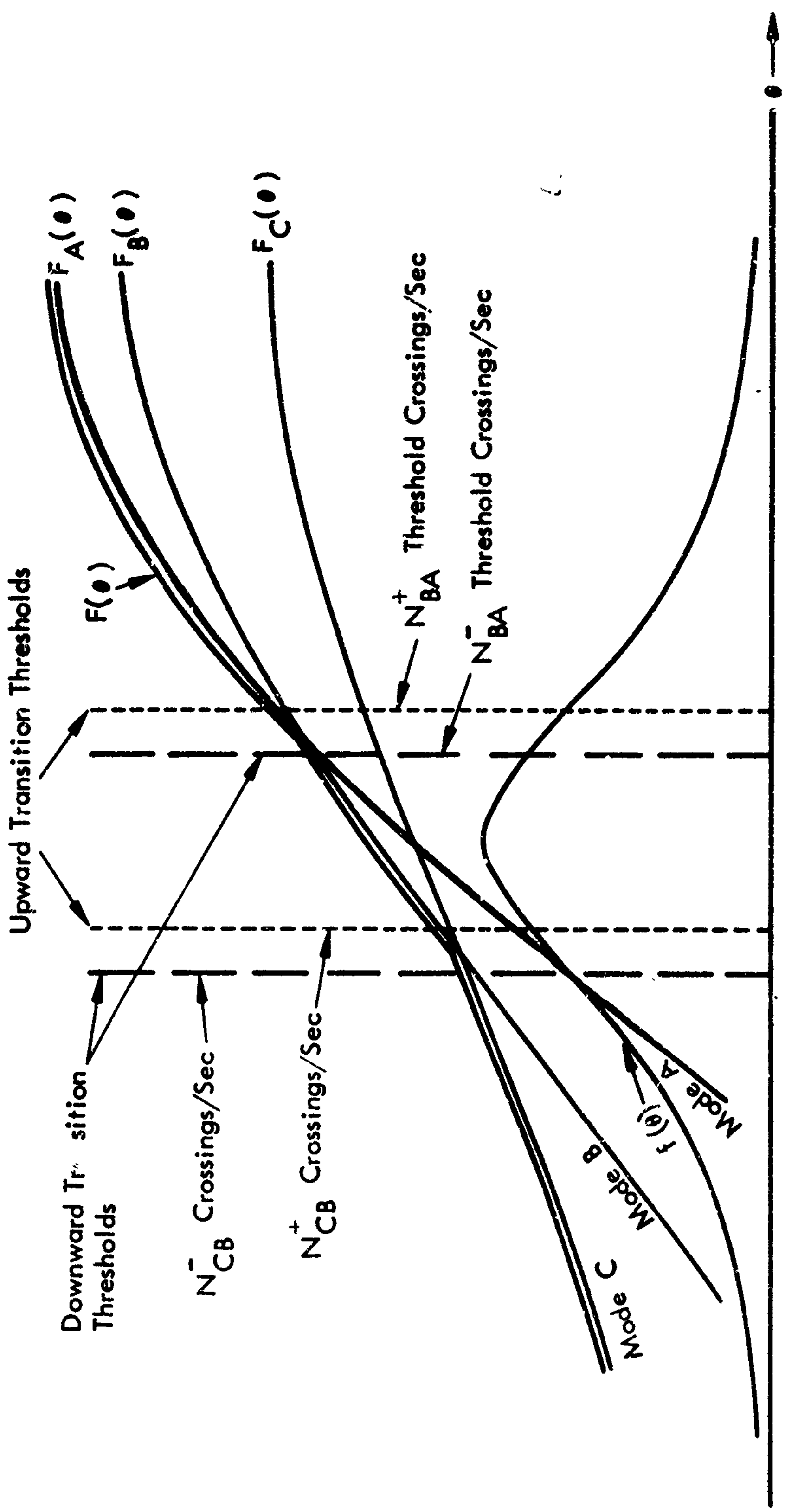


Figure 4-2. Illustrative Thresholds and Evaluation Functions

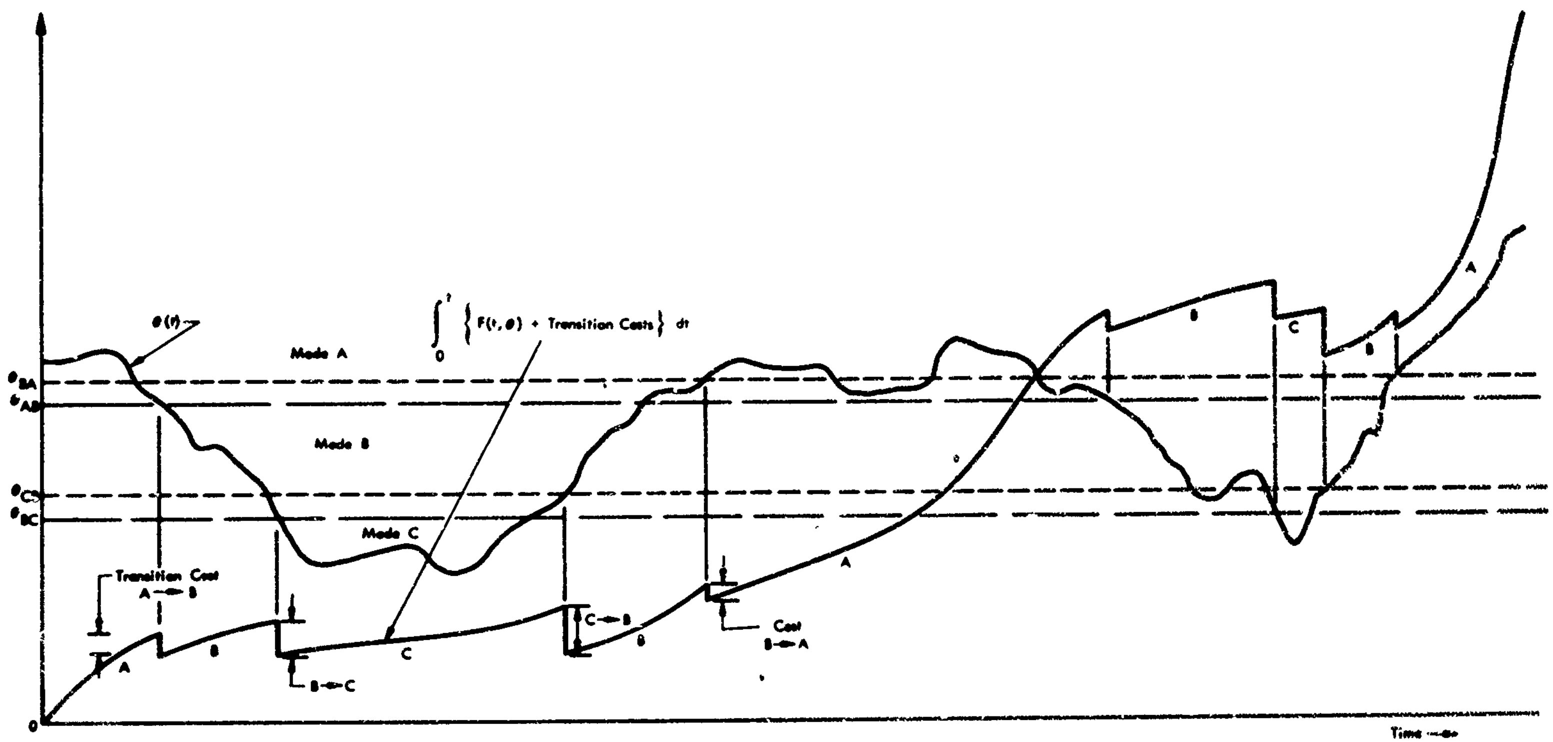


Figure 4-3. Hypothetical Evaluation Integral with Transition Costs

where r is the number of redundant bits in a coded block of n digits. If data is not encoded, $r = 0$. For a system which discards erroneous blocks of data, the throughput depends on the threshold variable (e.g., received signal power, signal-to-noise ratio) and can be represented as

$$R = F(\theta) = d\left(\frac{n-r}{n}\right) \log_2 m \,(1 - P_E(\theta))^n \qquad (4\text{-}7)$$

where $P_E(\theta)$ is the probability that a symbol will be received erroneously when the threshold variable has value θ.

The probability of bit error ρ, is given by

$$\rho = F(\theta) = P_E(\theta)\, \frac{1}{2}\, \frac{m}{m-1} \qquad (4\text{-}8)$$

where the correction factor takes into account that not all of an m-ary digit's information is lost when it is in error [4]. If the system has no error control, then $P_E(\theta)$ is the unconditional output bit error probability. If the system uses error correction, then $P_E(\theta)$ is the probability of uncorrected error for the particular code and threshold variable value θ, and ρ is the probability of an uncorrected bit error. Similarly, when error detection is used ρ and $P_E(\theta)$ represent probabilities of undetected error.

The transinformation or mutual information rate I bits per second combines the data rate and the error probability into a single number given by

$$I = F(\theta) = (n-r)\log_2 m \,\frac{1}{T}\left[1 + P_e \log_2 P_e + (1-P_e)\log_2(1-P_e)\right] \qquad (4\text{-}9)$$

where T is the symbol duration and $P_e = \frac{1}{2} \frac{m}{m-1} P_E$ is the equivalent binary error probability corresponding to the m-ary error probability P_E.

In an attempt to establish meaningful criteria for evaluating an adaptive communications system, an "evaluation function" has been proposed, along with a scheme for computing a performance index for each specific proposed system. The difficulty in this approach is that it is not clear how to define an evaluation function which is directly proportional to the "quality" of the system, without a good definition of "quality." Thus, arbitrarily specifying an evaluation function or a set of evaluation functions, one in effect sets the rules for design of the overall system. Because of the difficulty in specifying a set of evaluation functions, another approach is to evade the problem by leaving the "quality" ill-defined, and comparing systems based on sets of measures. One such scheme involves comparing systems on the basis of the pair (R, ρ), the throughput rate in bits per second, and uncorrected error rate in bits per bit. The problem of threshold selection is complicated by the fact that no linear ordering of the pairs of R and ρ exists, but once the thresholds are determined, the same integral expression can be applied yielding two performance indices, say I_R and I_ρ, with large I_R and small I_ρ being the desired goals. Constraints on minimum acceptable throughput rate and maximum tolerable error rate can be used to define permissible modes of operation for various ranges of the threshold variable.

4.1.4 Unsuitability of Mutual Information Alone as an Evaluation Function

Because it is desirable to formulate a performance index enabling a linear ordering of systems, a single measure of performance was sought which would incorporate the effects of errors, speed of transmission, etc. A seemingly natural choice for such a measure is the "mutual information" or "transinformation" rate, I (Equation 4-9), which has formal justification in the field of information theory. However, as can be seen from Figure 4-4, an error rate of 0.11 is required (on a per-bit basis) to reduce the mutual information by one-half of its zero-error figure. For an equivalent binary error probability of 0.01, the mutual information is reduced to 0.92 of the zero error rate; and for $P_E = 0.001$, to 0.99, approximately. Thus, the difficulty is that a very small change in I is associated with an intolerably large increase in error rate. Therefore, using this measure, a very large increase in error rate would be needed before it would be advantageous to operate in a slower speed, less-erroneous mode.

Figure 4-5 shows a set of approximate curves for 16, eight and two level FSK, with linear scales for S/N ratio and information rate. It is seen that the 16 level mode is best to the right of point β, eight levels are best for S/N between point α and point β, and two levels (binary) are best for the tiny fraction of the diagram to the left of point α. It should be noted that the distribution of S/N will be exponential for Rayleigh envelope. Even though the exponential distribution density heavily weights the small values of S/N, these values correspond to error rates not normally considered acceptable (e.g.,

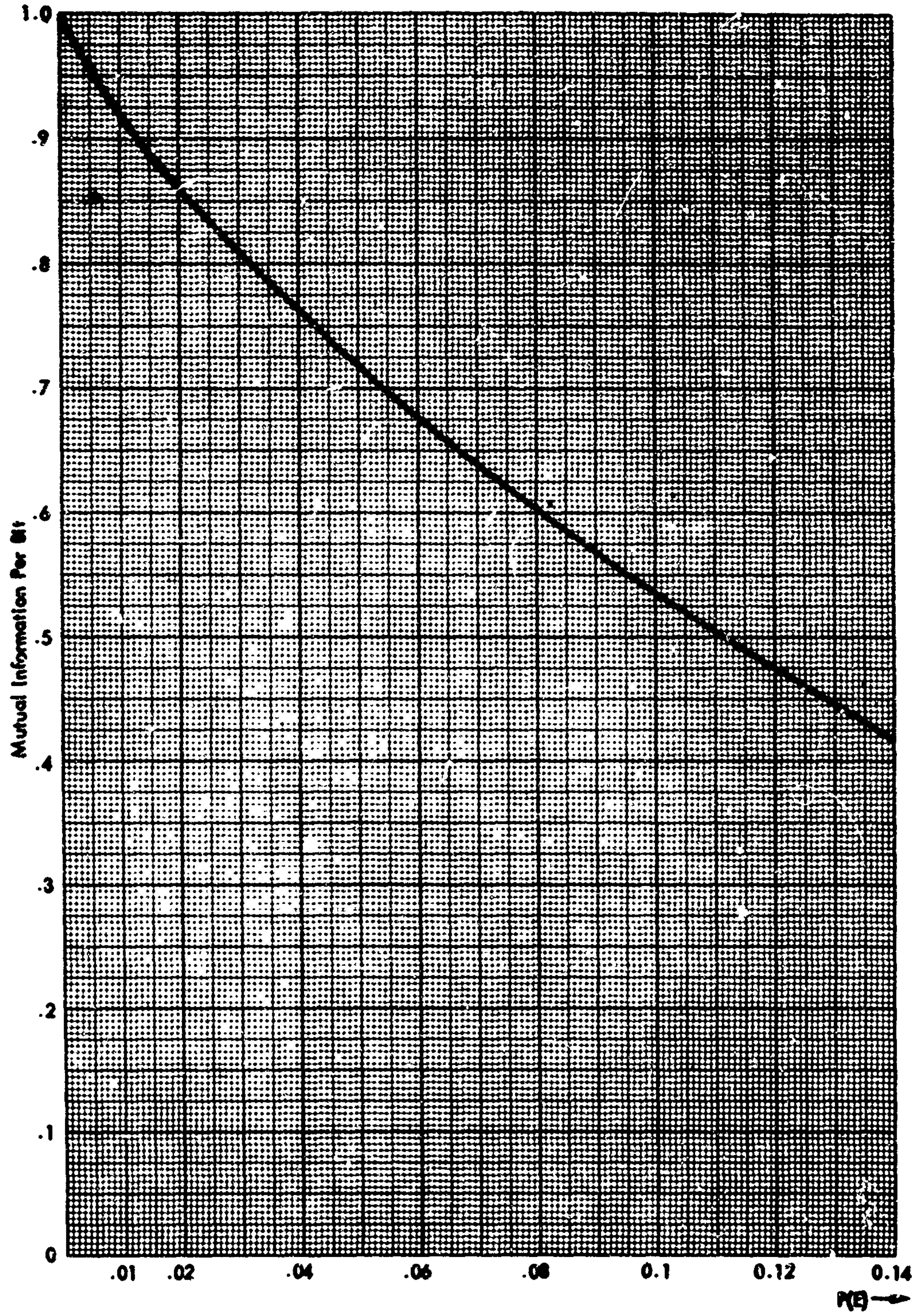


Figure 4-4. Mutual Information Per Bit as a Function of the Bit Error Probability in a Binary Symmetric Channel

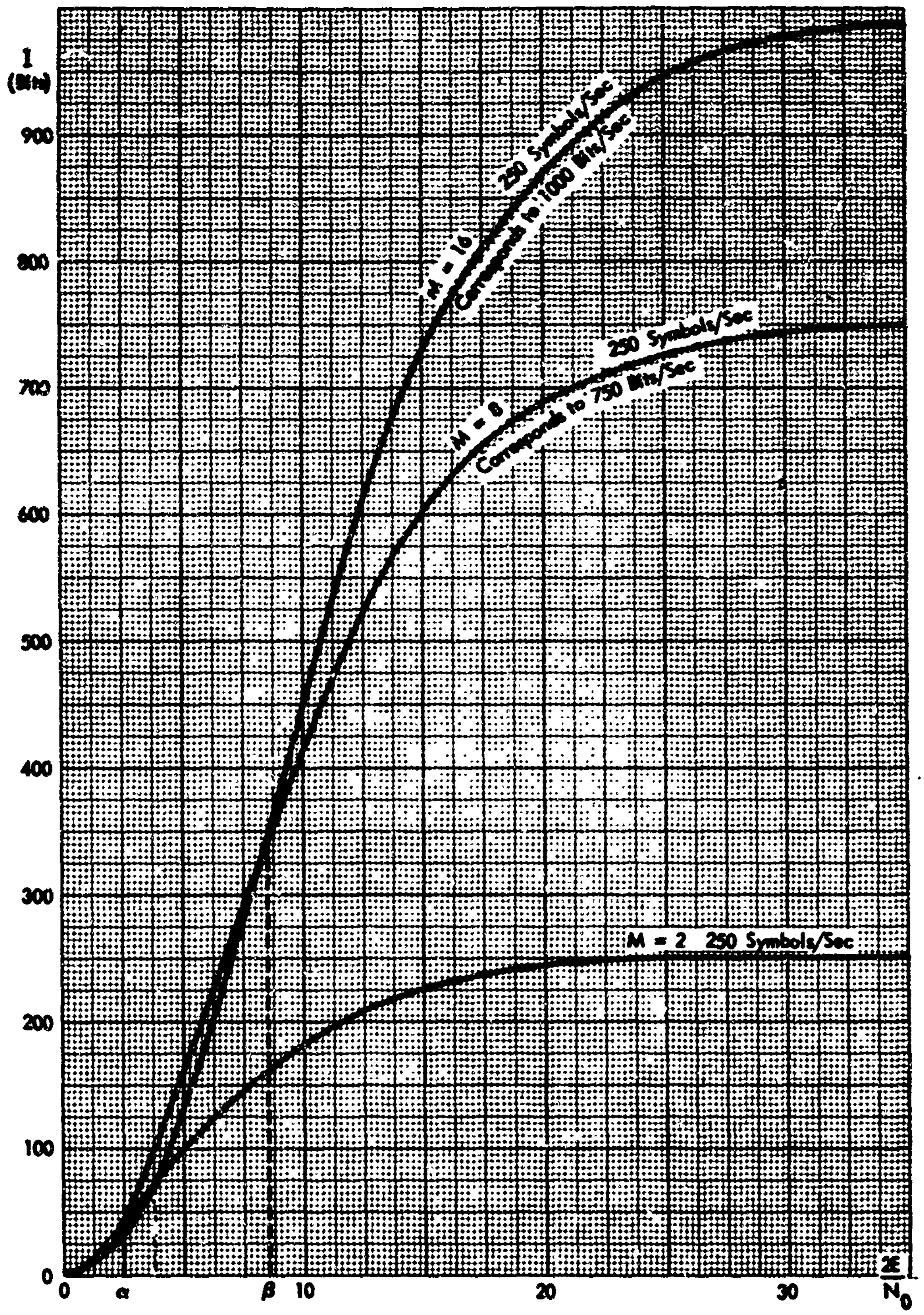


Figure 4-5. Relative Throughput with M-ary FSK

greater than 0.1). Thus, it is expected that a realistic adaptive scheme will make transitions at larger S/N ratios. This implies that a measure of performance different from mutual information is needed.

The curves in Figure 4-5 were obtained using Figure 4-4 for the information per binary digit as a function of the equivalent binary digit error probability. The equivalent binary digit error probability was obtained from the curves of Figure 4-6, which show symbol error probability vs $\frac{2E}{N_0}$, which is effectively a signal energy-to-noise power ratio. These curves are based on a paper by Nuttall [60]. The symbol error probabilities are converted to equivalent binary digit error probabilities using Equation (4-8).

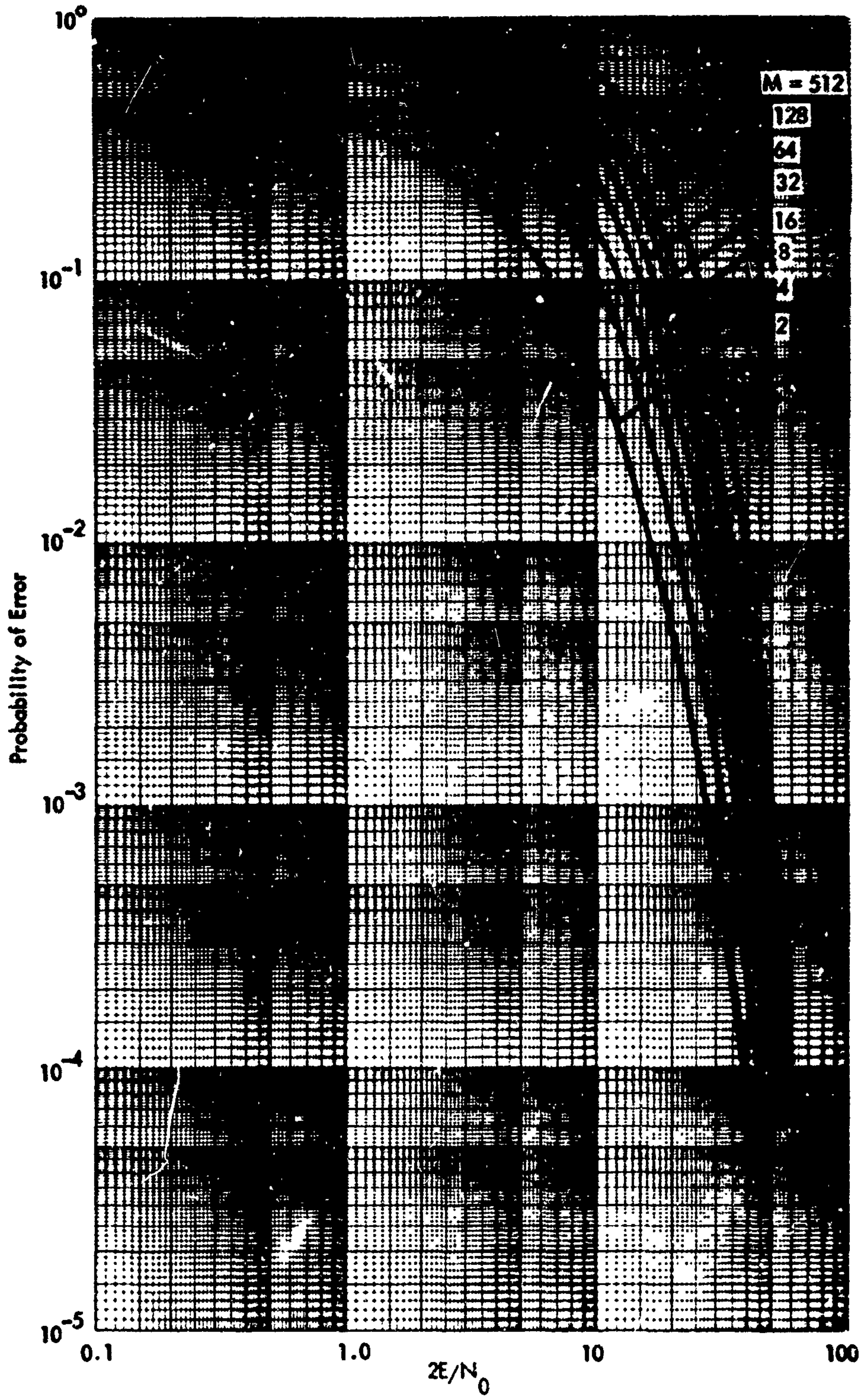


Figure 4-6. Error Probabilities for Phase Incoherent Maximum Likelihood Detection of Orthogonal M-ary Signals

255-nautical-mile path using 505 mc. Measurements were obtained over irregularly spaced runs throughout a two-month period, with the duration of each run, T_{int}, taken to be 15 minutes. To eliminate the variations in median signal levels from run to run, the signal level was expressed in db with respect to the median level of the signal during the particular run. They found that most of the runs resulted in an approximately Rayleigh distribution. As examples of their findings, fades at 505 mc, 15 db below the median signal level, and more than one second in duration occurred approximately 35 times per hour; 10 db fades lasting more than one second occurred approximately 70 times per hour. It is interesting to note that at 4090 mc the fast fades occured about eight times faster. Hourly median signal levels over a period of one year at 505 mc and 4090 mc on the 150-nautical-mile path were also measured. The measured hourly median signal levels were converted into the path losses in excess of the free space loss and expressed in terms of db. This is a reasonable way to normalize the hourly median signal levels, but one has to exercise care in choosing the antenna gains to compute the free space loss. As expected, the results are reasonably close to log-normal with standard deviations of 9 db and 7.5 db for the 505 mc and 4090 mc cases, respectively.

The variation in the median signal level is due mainly to seasonal variations. Less loss is incurred during the summer and fall in comparison to winter and spring. There is also evidence of diurnal variations, but these appear to be less significant than seasonal variations for the experiment under discussion.

For a 255-nautical-mile path, the hourly fading depth was measured for a carrier at 505 mc [18]. The measurement was made by transmitting 500 watts at 505 mc. A radio receiver incorporating a 20 kc narrowband IF amplifier was used. The receivers measured the magnitude of the signal power delivered by the receiving antenna. These measurements are summarized below.

Hourly Fading Depth (winter measurement) x db	Percent of Time Fading Depth is Less Than x db
7	1
10	10
11	20
15	65
18	90
22	99.5

For the same path and frequency, the number of fades per hour was related to the duration of the fades, as shown below.

Magnitude of Fade Below Hourly Median Value (db)	Fade Duration (sec)	Number of Fades per Hour with Longer Duration
15	1.0	32
	1.2	20
	1.5	10
	1.8	5
	2.0	3
10	2.0	30
	2.5	20
	4.0	10
	6.0	5
5	3.0	32
	4.8	20
	8.0	10
	16.0	5
	35.0	2

division multiplexed channels. In each channel only flat fading occurs. In their analysis, a pilot tone at f_c cps is received as

$$g_1(t)\, e^{j2\pi f_c t} \tag{4-10}$$

where $g_1(t)$ is a complex valued function with a normal distribution. If a pilot tone is transmitted at frequency f_c +F cps, the received signal is

$$h_1(t)\, e^{j\left[2\pi\ (f_c + F)t + \theta\right]} \;. \tag{4-11}$$

It is apparent that the channel state at f_c + F cps can be estimated on the basis of a pilot tone at f_c cps and the crosscorrelation between $g_1(t)$ and $h_1(t)$. Bello and Nelin worked with HF propagation, so to a good approximation $g_1(t)$ and $h_1(t)$ have a joint distribution which is gaussian.

4.2.2 Determination of Signal Level

The basic measurement is the determination of the incoming signal level. The input will consist of signal and noise, so the measured signal level will be the total input signal level. Under the assumption that the input noise is white gaussian with zero mean and relatively fixed noise power density the signal power can be obtained from:

$$\overline{(s(t)+n(t))^2} = \overline{s^2(t)} + \overline{2s(t)n(t)} + \overline{n^2(t)} \;. \tag{4-12}$$

Since the noise n(t) is independent from the signal s(t) and since $\overline{n(t)} = 0$, the above becomes:

$$\overline{(s(t) + n(t))^2} = \overline{s^2(t)} + \overline{n^2(t)} \tag{4-13}$$

whence

$$\overline{s^2(t)} = \overline{(s(t) + n(t))^2} - \overline{n^2(t)} \tag{4-14}$$

For an FSK system, this measurement can be performed at the output of the baseband circuit that forms the difference between the output of the detectors following the two band pass filters. If we denote the detector output signal by s(t) and the noise at the output of each of the detectors by $n_1(t)$ and $n_2(t)$, respectively, then when the signal is observed in the presence of the noise signal $n_1(t)$, the signal power at the output of the difference combiner is

$$\begin{aligned}\overline{(s(t) + n_1(t) - n_2(t))^2} = {} & \overline{s^2(t)} + \overline{n_1^2(t)} + \overline{n_2^2(t)} - \overline{2n_1(t)n_2(t)} \\ & + \overline{2s(t)n_1(t)} - \overline{2s(t)n_2(t)}\end{aligned} \tag{4-15}$$

When the signal s(t) is observed in the presence of noise signal $n_2(t)$, the corresponding signal power is

$$\begin{aligned}\overline{(n_1(t) - s(t) - n_2(t))^2} = {} & \overline{s^2(t)} + \overline{n_1^2(t)} + \overline{n_2^2(t)} - \overline{2n_1(t)n_2(t)} \\ & - \overline{2s(t)n_1(t)} + \overline{2s(t)n_2(t)}\ .\end{aligned} \tag{4-16}$$

The independence of the signal and noise terms was assumed. Obvious simplifications result if the noise has zero mean. The channel state can be estimated on the basis of these measurements.

When bi-phase modulation is used, a measurement at IF can be performed by using a square-law detector followed by filtering at twice the input signal frequency. During an information bit transmission, the phase angle due to modulating information is constant. The received signal in this case is narrow band gaussian noise which is expressible as

$$x_c(t) \cos \omega_c t + x_s(t) \sin \omega_c t \tag{4-17}$$

where $x_c(t)$, $x_s(t)$ are independent random variables with gaussian distribution and zero mean. Filtering the output of the square-law detector at frequency $2\,\omega_c$ one obtains

$$\frac{1}{2}\,(x_c^2(t)-x_s^2(t)) \cos 2\omega_c t + x_c(t)\, x_s(t) \sin 2\omega_c t \ . \tag{4-18}$$

The amplitude of the above signal is

$$\frac{x_c^2(t) + x_s^2(t)}{2} \ . \tag{4-19}$$

This is the average power of the received signal envelope. Square-law detection followed by low pass filtering to recover the dc term gives similar results.

The probability density function of $\sqrt{x_c^2(t) + x_s^2(t)}$ is Rayleigh, and a simple exercise reveals that the probability density function of $x_c^2(t) + x_s^2(t)$ is

$$\frac{1}{2\sigma^2} \exp \left[- \frac{x_c^2(t) + x_s^2(t)}{2\sigma^2} \right] \tag{4-20}$$

where σ is the variance of the narrow band gaussian process.

The basic measurement is made to estimate the average received signal power, which in turn provides information concerning the time varying channel state. The general functional relation between the per digit error rate, received signal energy per bit, and received noise power density is

$$P = f\left(\frac{E}{N_0}\right) \tag{4-21}$$

where E is the energy received per bit and N_0 the noise power density. Examples of this relationship are given in subsection 3.4. Since

$$E = \overline{s^2(t)}T, \tag{4-22}$$

where T is the information bit duration, the error rate, P, can be obtained after the reception of each information bit. If the system performance is specified in terms of the error rate, the estimated error rate based on the measurement of average received signal power and Equations (4-12), (4-21), and (4-22) provides a measure of how close to the desired error rate the system operates. If the desired and estimated error rates are denoted by

P_{des} and $P_{est}(t)$ respectively, the system adaptation to the channel state must minimize $P_{des} - P_{est}(t)$.

That the received signal level, after the multipath signals from the previous bit duration have subsided, can be used as a basis for adapting to the varying channel states was demonstrated by Lieberman [2]. He used the received signal-to-noise ratio to adapt his communication system, and points out that when the signal-to-noise ratio is sufficiently high, this quantity is proportional to the channel capacity.

4.2.3 Use of Error Control Techniques

In addition to directly measuring channel parameters, it is also possible to use error detecting or error correcting codes to assess the channel. This differs from measurements of the other parameters discussed here in that it is an after-the-fact channel measurement, i.e., the received signal has been detected, bit decisions made, and receiver noise has been introduced. This is, however, not necessarily undesirable. User requirements are generally stated in terms of quality of data at the system output and error control measurements evaluate the quality of the data output directly. Thus, although not as good as a channel evaluator, error control is a superior data quality monitor. Possibly a combination of diversity measurement and error control measurement is optimum.

Certainly if the adaptive system is already an error control system, the existing coding may be used to assess the channel rather than introducing additional equipment. Criteria such as the number of errors in consecutive blocks, consecutive blocks containing errors, and consecutive error-free blocks were used to make mode change decisions in the simulation of the fixed block length, variable redundancy and variable block length, fixed redundancy techniques (see Section 6). On the other hand, the inherent delay associated with the generation of error information may limit the usefulness of this technique. If any attempt is made to extend the redundancy over long blocks, or if the information bits are distributed and not taken sequentially, a further delay will be incurred in obtaining the information pertaining to the number of bits in error. It should also be noted that this is a coarse block-by-block measurement as compared with the continuous measurement associated with the other techniques.

4.3 DELAYS

In reality, it is not possible to adapt a system the instant a change in the environment occurs. Indeed, several types of delays are inherent in any adaptive communication system.

First there is a delay in sensing that the channel state has changed. The two main techniques for monitoring the channel (subsection 4.2) are, in effect, signal strength measurement and error estimates from coding. The former requires some time for processing and integration in order that a meaningful average value is obtained rather than some extremely short-lived dip observed in a brief sample. The latter requires that a complete coded block of symbols be received and processed before an indication is available regarding the presence or absence of errors.

Second, there is a delay in conveying the channel state information or mode change request to the transmitter. This delay is the sum of two delays. One is the propagation delay between the terminals. The other, and generally larger, is a delay caused by constraints on the adaptive control logic (subsection 4.5). It is necessary that all messages follow a fixed format. Consequently channel and mode information, which is really only a minor part of the data being sent, cannot be sent at will, but must be sent only at particular times. Thus, there is a waiting period to inform the other terminal.

Third, there is a delay in reacting to the receipt of new channel state information or a mode change request. Here again, format restrictions and synchronization problems limit the specific times when a terminal may change, and there will be a delay while the terminal waits for the next possible time to change modes.

The delay problem restricts those situations in which adaptive techniques can effect an improvement in performance in a realistic environment. A further discussion of this aspect appears in subsection 4.4. For those situations in which the delay problem does not preclude the use of adaptive techniques, the extent of improvement is reduced. In Section 5, actual systems are analyzed, and the effects of delays on performance are expressed both in formulas (subsections 5.3 to 5.5) and quantitatively (subsection 5.6).

4.4 IMPROVABLE SITUATIONS

As previously shown, a major limiting factor to effective adaptive communications is the delay from the time the channel changes to the time the system can adapt. It is immediately evident that channels subject to random errors cannot be improved by the techniques covered in this report—with the exception of those which repeat erroneous data in some form. The main value of the techniques not currently in use will be in fading channels, where the degraded channel state persists for a reasonable length of time. Here the adaptive change can take effect and give improvement to compensate for the additional cost.

Not all fading can be overcome by adaptive techniques (other than diversity). For example, some Rayleigh fast fading changes so rapidly that adaptive control is infeasible. This is not at all due to an expected limit on the speed of automatic equipment, but represents a consideration of the intrinsic delays in communicating adaptive control information discussed in subsection 4.3.

In computing the average delay after a threshold is crossed until transmissions can be received in a new mode, an important, perhaps controlling, factor is the length of the block structure imposed on the data. This block structure is required to provide a vehicle for control messages and switchovers. The importance of the block length derives from the fact that the threshold crossings will occur at random within some block interval. Therefore, the average delay before the next possible service message transmitted in the reverse direction will be $\frac{1}{2}$ T_b, where T_b is the block duration in seconds. Assume that the service message starts T_s seconds before the end of a block. Then the minimum delay from the time a threshold is crossed at one terminal until the other terminal finishes receiving the service message is $(T_s + T_p)$, where T_p is the one-way propagation time. The corresponding maximum delay is $(T_s + T_p + T_b)$. Typical values might be $T_p = 1$ ms (5 sec/mile for 200 mile link), $T_s = 20$ ms (10 bits at 500 bits/sec), and $T_b = 200$ ms (100 bits at 500 bits/sec). Thus, the delay before a service message is received at the transmitting terminal will be uniformly distributed, with mean $(2T_s + 2T_p + T_b)/2$ which, for the values given as an example above, would yield an average delay of 121 ms. (This assumes no confusion or error in the transmission of the service message.) There will be an additional constant delay T_d (typically half a block) after the service message is received before transmission in the new mode can begin. Thus, for the present example the overall average time from threshold crossing to reception of signals in the new mode is about 221 ms. To illustrate the infeasibility of adapting to Rayleigh fast fading, consider the average duration of fades below a given level and the average number of fades per second below a given level, as given for tropospheric propagation in Reference [62]. In that reference, it is stated that for fading rates (i.e., threshold crossings at the median level) greater than about 12 per minute for 4110 mc and 8 per minute for 460 mc, the envelopes of the signals are Rayleigh-distributed. The fading rates depend on wind velocities and turbulence in the reflection volume and are so highly variable that it is useless to give a single "typical" fading rate. However, the median fading rates for a 460-mc troposcatter link show a diurnal variation of the median fading rate from about 7 to 19 fades per minute. Furthermore, for higher frequencies, the fade rate will be greater by a factor which usually exceeds the ratio of frequencies. As an illustration, assume a median fading rate of 12 crossings per second. Then, assuming a two-channel combiner and a threshold at 0 db, the average crossing rate in a single direction would be about 8.4 crossings per second with an average duration of the signal below the 0-db level of about 220 ms [62]. Since the average occupancy time in a mode is thus comparable in duration to the average delay in making a transition, the possibility of usefully adapting to these fades is virtually ruled out.

Thus, it is only the slower fading phenomena (Rayleigh or log-normal) to which we can adapt. As seen in Section 3, tropospheric scatter, ionospheric scatter, and ionospheric reflection all have slow fades as well as rapid fades. Although land lines have long disturbances, these are so effectively handled by high efficiency, high reliability, error detection-retransmission systems that they present no problem and offer little room for improvement [63].

Quantitative results on the degree of improvement in a wide variety of channels are given in subsection 5.6 for variable modulation parameter techniques and in subsections 6.2 and 6.3 for variable error control techniques.

4.5 CONTROL PROBLEMS

The control problems associated with adaptive data communication systems are critical to the proper performance of such systems. They are generally dismissed lightly as either trivial or obvious. Nevertheless, subtle variations and seemingly inconsequential short-comings can render a potentially useful system ineffective or inoperative.

The main function of control information is to enable the receiving terminal to convey to the transmitting terminal information as to what changes should be made in the mode of transmission and when these changes should be made. The major problem associated with this is based on the fact that this control information is subject to errors in transmission which may or may not be detected by the recipient. There must be a well defined course of action to cover all eventualities lest situations arise in which neither terminal knows what the other is doing or what it should be doing itself, resulting in chaos that prevents any meaningful transmission of data.

If the recipient of the control information recognizes that a particular message contains an error, he can reject it or convey his uncertainty to the sender. If he does not recognize that a particular control message is wrong and adapts erroneously, the receiver may lose synchronization and may also be unable to distinguish among redundancy, data, and control information. Even when the recipient can identify faulty control information, the sender will still be unaware that it was faulty and will be expecting an adaptive change. When this change does not occur, sync may be lost. Furthermore, attempts to convey this "faulty receipt" back to the sender are themselves subject to error, and so on.

What is needed, then, is a set of operating rules and procedures to make it unlikely that the terminals lose awareness of what they should be doing or of what the other is doing. In addition, provision must be made for the system to recover from the infrequent, but never completely avoidable, breakdowns. This is the purpose of the adaptive control logic.

In designing such a logic, it is desirable to make it as fast-acting as possible so that excessive or intolerable delays are not introduced. Such delays would render the adaptive action useless when it finally occurred (subsection 4.3). Furthermore, it should be as efficient as possible. Each bit of control information sent replaces a bit of data. Extra controls must not lower data throughput more than the adaptive action will increase it.

Various control logics, both old and new, are discussed in subsection 4.6.

4.6 ADAPTIVE CONTROL LOGICS

Every adaptive system needs a control logic which unequivocally defines a course of action for every possible eventuality. For retransmission systems, several such logics have been developed. These logics, together with some improvements, are discussed below. In addition, a new logic for nonretransmission systems was developed and is explained in this section.

4.6.1 Retransmission Logics

The control logic for a retransmission system must indicate when to retransmit, what to retransmit, what to accept, and what to do when ambiguous control messages are received. It must also provide a means for system recovery from undetectable errors.

4.6.1.1 Fixed Format Systems

The most common retransmission systems always transmit both data and control information in the same format, either separately or together. There are three basic types: Idle-RQ, Simple-RQ, and Dual-RQ. Virtually every retransmission system design is one of these or a minor variation of one of these. The reader is referred to Reference [13] for a detailed description and analysis of these control logics.

4.6.1.2 Variable Format Systems

It is also possible in a retransmission system to vary the block length or the amount of redundancy in a block. If the block length is to be varied, more control information and operating procedures than required for fixed block retransmission will be needed. Such a control logic, adaptive Dual-RQ, was developed by Corr and Frey [1]. It uses two control bits and an elaborate set of rules for switching, printing, and accounting. Although it appears complex, it is the simplest design known that allows continuous operation without loss of data.

If the block length is to be kept fixed, but the amount of redundancy varied, then the regular Dual-RQ control logic suffices except for a problem of data alignment. This problem is the result of two facts:

1. On a block-by-block basis, the retransmitted blocks do not contain the same data bits as the original blocks whenever a mode change accompanies the retransmission, for the new blocks have the same overall length but different redundancy length and hence different data length.
2. The retransmission logic does not make it clear to the receiving terminal where regular transmissions end and repeat transmissions begin.

In fixed-block, fixed-redundancy Dual-RQ this problem does not arise since the proper data is automatically sent to the destination, regardless of which blocks are discarded. In contrast, the number of data bits discarded in the variable system varies as the modes change.

Although further analysis is required to verify that the data stream can be realigned in every conceivable case, it appears that, since precise information concerning the number of data bits in any given block is contained in the identity of the code polynomial for that mode, the receiver can always determine the number of data bits in any correct block. Hence, the alignment of the data and mode changes.

4.6.1.3 Improvements in the Design of Retransmission Systems

Dual-RQ type retransmission logics are probably the most satisfactory and widely used in practice today. The basic version described by Reiffen, Schmidt and Yudkin [63] represents a significant improvement over van Duuren's original design [64]. The basic version's complete behavior was analyzed and described in 1964 [13]. Two modifications of the basic version which lead to simpler implementation and improved performance are described in the following paragraphs.

Briefly, the Basic Dual-RQ is a full duplex system in which both terminals send message blocks of the same size. These blocks contain a service bit reporting whether the most recently received block contained errors. Each block is protected by a strong error detecting code. When an erroneous block is received by a terminal, or when word that the other terminal received an erroneous block is received, the terminal goes back two blocks retransmits these, and then continues (either retransmitting again or advancing). At the same time, two incoming blocks are discarded. A terminal state diagram for the Basic Dual-RQ is given in Figure 4-7. The matrix of transition probabilities between message states as defined in Reference [13] is given in Figure 4-8, and the matrix of message acceptance probabilities is given in Figure 4-9, where $d = P_d$ is the probability that a block will be detected in error, $u = P_u$ is the probability that a block will contain an undetected error, $a = P_\alpha$ is the probability that the service bit in a block containing an undetected error has been changed, and $SX = P_{SX}$ is the probability of going from state S to state X.

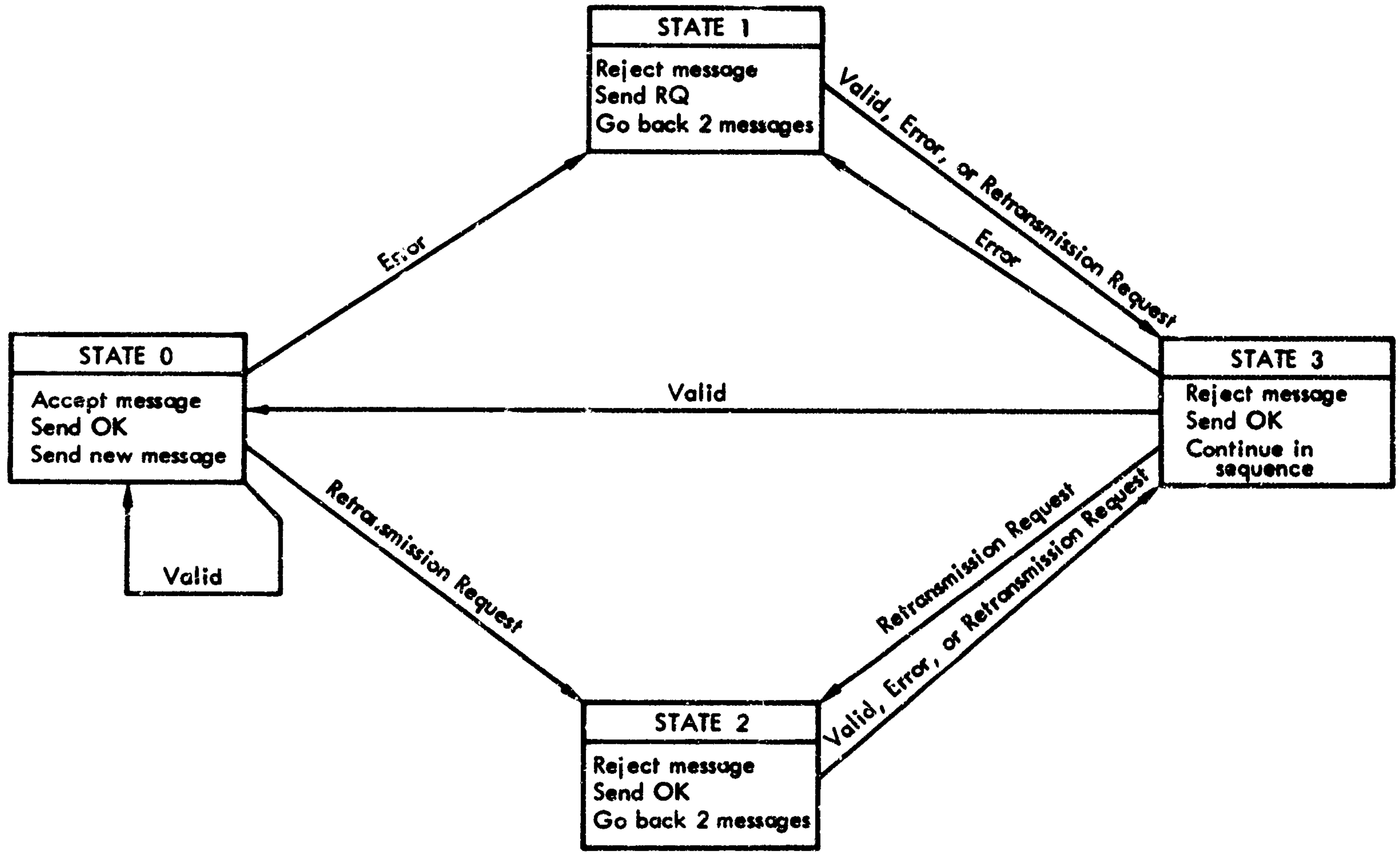


Figure 4-7. Terminal State Diagram—Basic Dual-RQ

	S	S'	A	A'	B	B'	C	D
S	0	0	$d + (ua)^2$	$d(1-d-ua)$	$(1-d-ua)^2$	0	$ua(1-ua)$	$ua(1-d-ua)$
S'	0	0	0	d	$1-d-ua$	0	0	ua
A	0	0	$d + (ua)^2$	$d(1-d-ua)$	$(1-d-ua)^2$	0	$ua(1-ua)$	$ua(1-d-ua)$
A'	0	0	$ua\delta A$	$ua\delta A' + d(1-ua)$	$ua\delta B + (1-ua)(1-d-ua)$	0	$ua\delta C$	$ua\delta D + ua(1-ua)$
B	$(1-d-ua)^2$	$ua(1-ua)$	0	0	$ua(1-ua)(1-d-ua) + d + (ua)^2$	$\left[d + (ua)^2\right](1-d-ua)$	0	0
B'	$1-d-ua$	0	0	0	$ua(1-ua)$	$d + (ua)^2$	0	0
C	$1-d-ua$	0	0	0	$ua(1-ua)$	0	$d(1-ua) + ua(d + ua)$	0
D	0	0	$(1-d-ua)\delta A$	$(1-d-ua)\delta A' + d(d + ua)$	$(1-d-ua)\delta B + (d + ua)(1-d-ua)$	0	$(1-d-ua)\delta C$	$(1-d-ua)\delta D + ua(d + ua)$

Figure 4-8. Matrix of Transition Probabilities for Basic Dual-RQ

	R*	A	A'	B	B'	C	D	U	V	W
R*	0	$\frac{S}{S+S'}\left[d+(ua)^2\right]$	$\frac{S}{S+S'}\left[d(1-d-ua)\right] + \frac{S'}{S+S'}(d)$	$\frac{S}{S+S'}(1-d-ua)^2 + \frac{S'}{S+S'}(1-d-ua)$	0	$\frac{S}{S+S'}(ua)(1-ua)$	$\frac{S}{S+S'}(ua)(1-d-ua) + \frac{S}{S+S'}(ua)$	0	0	0
A	0	$d+(ua)^2$	$d(1-d-ua)$	$(1-d-ua)^2$	0	$ua(1-ua)$	$ua(1-d-ua)$	0	0	0
A'	0	$ua\beta A$	$ua\beta A' + d(1-ua)$	$ua\beta B + (1-ua)(1-d-ua)$	0	$ua\beta C$	$ua\beta D + ua(1-ua)$	0	0	0
B	0	0	0	$u(1-a)(ua)(1-ua) + d + (ua)^2$	$u(1-a)\left[d+(ua)^2\right]$	0	0	$u(1-a)(1-d-ua)$	$ua(1-ua)$	$1-d-u$
B'	0	0	0	$ua(1-ua)$	$d+(ua)^2$	0	0	0	$1-d-ua$	0
C	0	0	0	$ua(1-ua)$	0	$d(1-ua) + ua(d+ua)$	0	0	$1-d-ua$	0
D	0	$u(1-a)\beta A$	$u(1-a)\beta A' + d(d+ua)$	$u(1-a)\beta B + (d+ua)(1-d-ua)$	0	$u(1-a)\beta C$	$u(1-a)\beta D + ua(d+ua)$	0	0	$1-d-u$
U	0	0	0	0	0	0	0	1	0	0
V	0	0	0	0	0	0	0	0	1	0
W	0	0	0	0	0	0	0	0	0	1

Figure 4-9. Matrix of Outcome Probabilities for Basic Dual-RQ

4.6.1.3.1 Revised Dual-RQ

It was observed that terminal states 1 and 2 perform identically except for the choice of the outgoing service bit. However, the message containing the OK from terminal state 2 is not examined by the opposite terminal when it gets there, but is automatically discarded because of the previously received erroneous blocks. Hence, an RQ might just as well be in the block. If this is done, the transmitter need not distinguish between erroneous blocks and RQ-bearing blocks, and terminal states 1 and 2 become identical in all respects. This results in the simplified terminal state diagram shown in Figure 4-10.

This Revised Dual-RQ logic behaves exactly as the Basic Dual-RQ logic in the absence of undetected errors. In the event of undetected errors changing a confirmation to a retransmission request, the Basic Dual-RQ repeats messages without the opposite terminal's knowledge; in the Revised Dual-RQ the opposite terminal is notified and goes into a repetition cycle. Thus the Basic Dual-RQ discards messages that will not be repeated, whereas the Revised Dual-RQ repeats these messages. In terms of message states, State D no longer exists. The transition and outcome matrices are given for the Revised Dual-RQ in Figures 4-11 and 4-12, respectively.

This simplification in terminal design is accompanied by an increase in reliability (the wrongly discarded messages are retransmitted) and a decrease in throughput (more retransmissions). A computer examination of the comparative performance of the basic and revised logics was made for all block lengths less than 1024 which offered improved reliability over uncoded

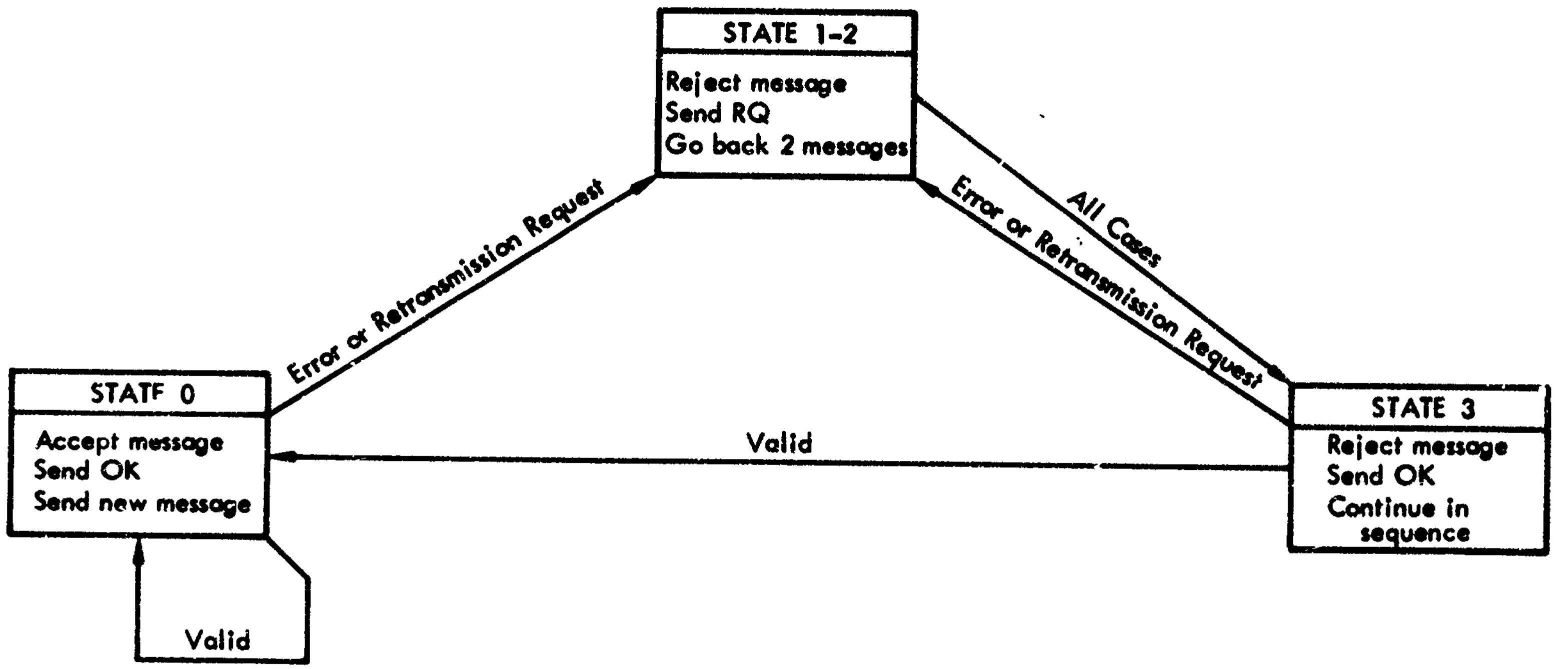


Figure 4-10. Terminal State Diagram—Revised Dual RQ

	S	S'	A	A'	B	B'	C
S	0	0	(d + ua)(1-ua)	(1-d-ua)(d + ua)	$(1\text{-}d\text{-}ua)^2$	0	$(ua)^2$ + dua
S'	0	0	0	d + ua	1-d-ua	0	0
A	0	0	(d + ua)(1-ua)	(1-d-ua)(d + ua)	$(1\text{-}d\text{-}ua)^2$	0	$(ua)^2$ + dua
A'	0	0	uaSA	uaSA' + (1-ua)(d + ua)	uaSB + (1-ua)(1-d-ua)	0	uaSC
B	$(1\text{-}d\text{-}ua)^2$	(d + ua)ua	0	0	(1-d-ua)(d + ua)(ua) + (d + ua)(1-ua)	(1-d-ua) (d + ua) (1-ua)	0
B'	1-d-ua	0	0	0	(d + ua)ua	(d + ua)(1-ua)	0
C	1-d-ua	0	0	0	(d + ua)ua	0	(d + ua)(1-ua)

Figure 4-11. Matrix of Transition Probabilities for Revised Dual-RQ

	S°	A	A'	B	B'	C	U	V	W
S°	0	$\frac{S}{S+S'}(d+ua)(1-ua)$	$\frac{S}{S+S'}(1-d-ua)(d+ua)$ $+$ $\frac{S'}{S+S'}(d+ua)$	$\frac{S}{S+S'}(1-d-ua)^2$ $+$ $\frac{S'}{S+S'}(1-d-ua)$	0	$\frac{S}{S+S'}(ua)^2+dua$	0	0	0
A	0	$(d+ua)(1-ua)$	$(1-d-ua)(d+ua)$	$(1-d-ua)^2$	0	$(ua)^2+dua$	0	0	0
A	0	uaAA	uaAA' $+$ $(1-ua)(d+ua)$	uaAB $+$ $(1-ua)(1-d-ua)$	0	uaAC	0	0	0
B	0	0	0	$u(1-a)(d+ua)ua$ $+$ $(d+ua)(1-ua)$	$u(1-a)(d+ua)(1-ua)$	0	$u(1-a)(1-d-ua)$	$(d+ua)ua$	$1-d-u$
B'	0	0	0	$(d+ua)ua$	$(d+ua)(1-ua)$	0	0	$1-d-ua$	0
C	0	0	0	$(d+ua)ua$	0	$(d+ua)(1-ua)$	0	$1-d-ua$	0
U	0	0	0	0	0	0	1	0	0
V	0	0	0	0	0	0	0	1	0
W	0	0	0	0	0	0	0	0	1

Figure 4-12. Matrix of Outcome Probabilities for Revised Dual-RQ

systems at a reasonable throughput rate. Both independent and dependent error channels were considered. For all dependent error cases and virtually all independent error cases, it was found that

$$\frac{\text{Revised Dual-RQ throughput}}{\text{Basic Dual-RQ throughput}} = 1.0000$$

In the few cases with block lengths 31 or less where this was not true, the ratio never went below 0.9990. Hence there is no significant decrease in throughput.

The examination of the uncorrected error rate ρ showed that, independent of the channel error probabilities, the following holds:

$$1.0000 \leq \frac{\text{Basic Dual-RQ } \rho}{\text{Revised Dual-RQ } \rho} \leq 3.000$$

The variation of the ratio with block length and code capability is shown in Figures 4-13 and 4-14 for dependent and independent errors, respectively. From these figures it can be seen that there is a small, but observable, improvement in reliability with the Revised Dual-RQ. Its main advantage remains, however, with the implementation.

4.6.1.3.2 No-Service Dual-RQ

Since repeat requests and errors are treated identically, it was observed that a special bit to request repeats is not necessary—many errors could be intentionally introduced in the outgoing block with same end result. The logic

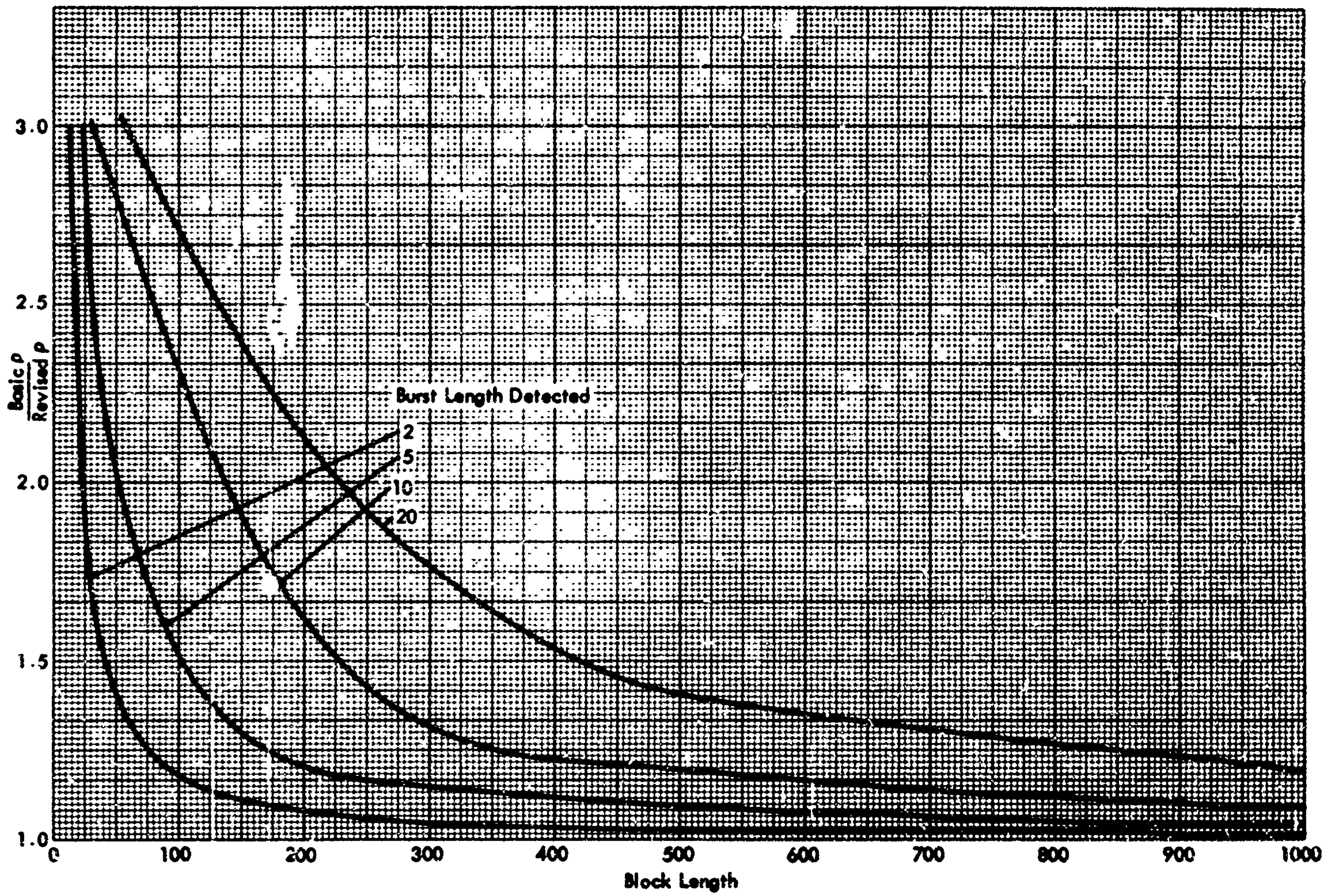


Figure 4-13. Reliability Improvement with Revised Dual-RQ, Burst Errors

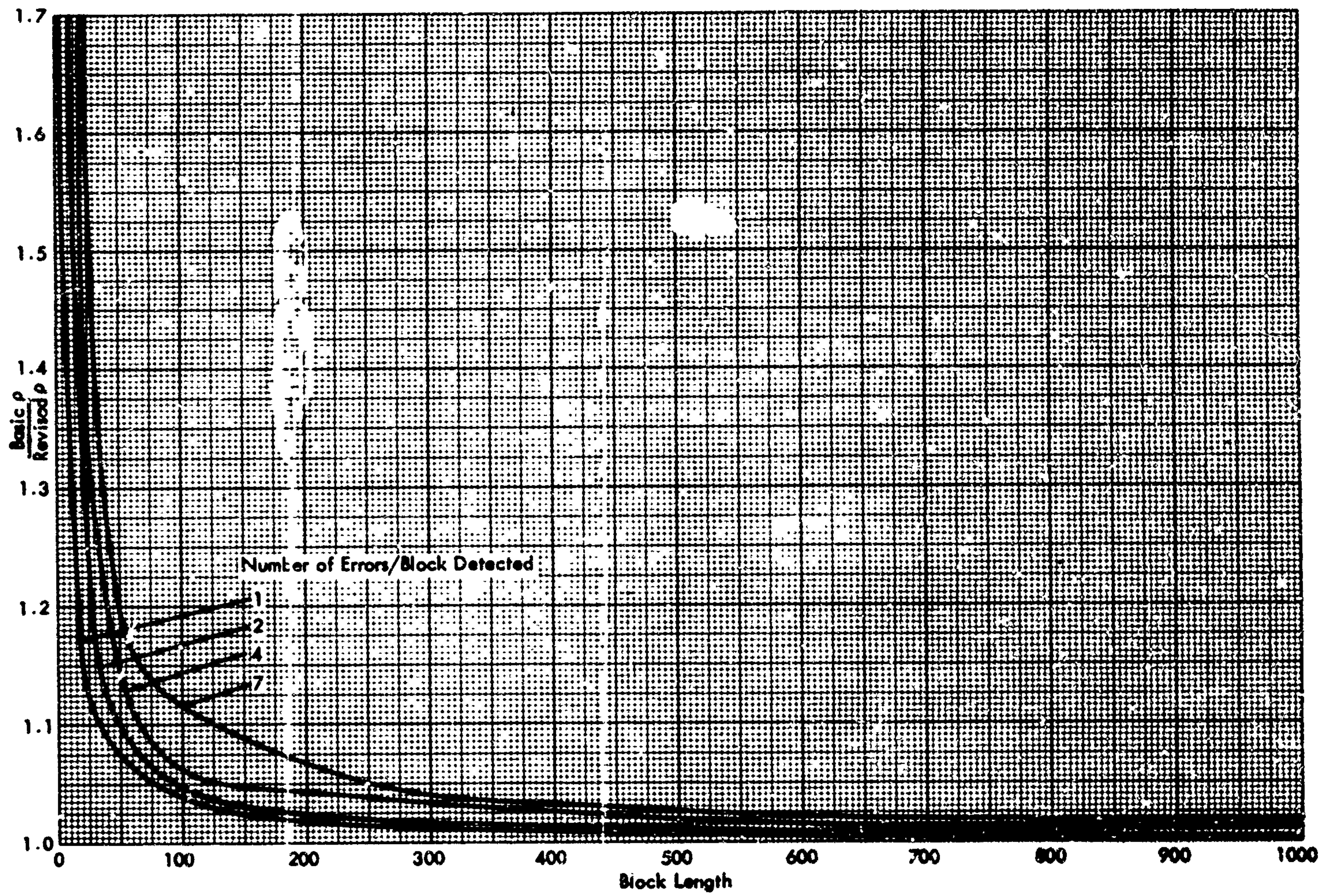


Figure 4-14. Reliability Improvement with Revised Dual-RQ, Independent Errors

with this modification was named the No-Service Dual-RQ. Its message states, transition and outcome probabilities were analyzed and are shown in Figures 4-15 and 4-16, respectively. They are the same as for the Revised Dual-RQ except that P_α, the probability that the service bit will be changed in a message containing an undetected error, is not defined. These terms are replaced by 1 for RQ $\rightarrow$ OK changes and 0 for OK $\rightarrow$ RQ changes. The logic was then evaluated for the same codes and channels as the two other logics. It was assumed that the undetected block error probability for the intentionally invalid blocks would be the same as that calculated for intentionally valid blocks.

An increase in throughput (a data bit is sent instead of the service bit of approximately $\frac{1}{k}$, where k is the number of data bits per block, was observed as expected. This is accompanied by a decrease in reliability (undetected errors are more likely) of one to three percent for independent error channels and one to 50 percent for dependent errors. The latter are independent of channel statistics, but vary according to block length and code strength as shown in Figure 4-17.

4.6.1.3.3 Conclusions

The Revised Dual-RQ is superior to the Basic Dual-RQ in implementation and reliability and equal in throughput. The No-Service Dual-RQ offers more throughput coupled with less reliability. This tradeoff must be evaluated by the individual user in the light of the problems introduced by the intentional

	S	S'	A	A'	B	B'	C
S	0	0	$d(1-u)$	$(1-d)d$	$(1-d)^2$	0	du
S'	0	0	0	d	$1-d$	0	0
A	0	0	$d(1-u)$	$(1-d)d$	$(1-d)^2$	0	du
A'	0	0	uSA	$uSA' + (1-u)d$	$uSB + (1-u)(1-d)$	0	uSC
B	$(1-d)^2$	du	0	0	$d(1-u) + (1-d)du$	$(1-d)(d)(1-u)$	0
B'	$1-d$	0	0	0	du	$d(1-u)$	0
C	$1-d$	0	0	0	du	$d(1-u)$	0

Figure 4-15. Matrix of Transition Probabilities for No-Service Dual-RQ

	S*	A	A'	B	B'	C	U	V	W
S*	0	$\frac{S}{S+S'}(d-du)$	$\frac{S}{S+S'}(d-d^2)$ + $\frac{S'}{S+S'}(d)$	$\frac{S}{S+S'}(1-d)^2$ + $\frac{S'}{S+S'}(1-d)$	0	$\frac{S}{S+S'}(du)$	0	0	0
A	0	$d(1-u)$	$(1-d)d$	$(1-d)^2$	0	du	0	0	0
A'	0	$ud(1-u)$	$ud(1-d) + (1-u)d$	$u(1-d)^2 + (1-u)(1-d)$	0	udu	0	0	0
B	0	0	0	$d(1-u) + udu$	$ud(1-u)$	0	$u(1-d)$	du	$1-d-u$
B'	0	0	0	du	$d(1-u)$	0	0	$1-d$	0
C	0	0	0	du	$d(1-u)$	0	0	$1-d$	0
U	0	0	0	0	0	0	1	0	0
V	0	0	0	0	0	0	0	1	0
W	0	0	0	0	0	0	0	0	1

Figure 4-16. Matrix of Outcome Probabilities for No-Service Dual-RQ

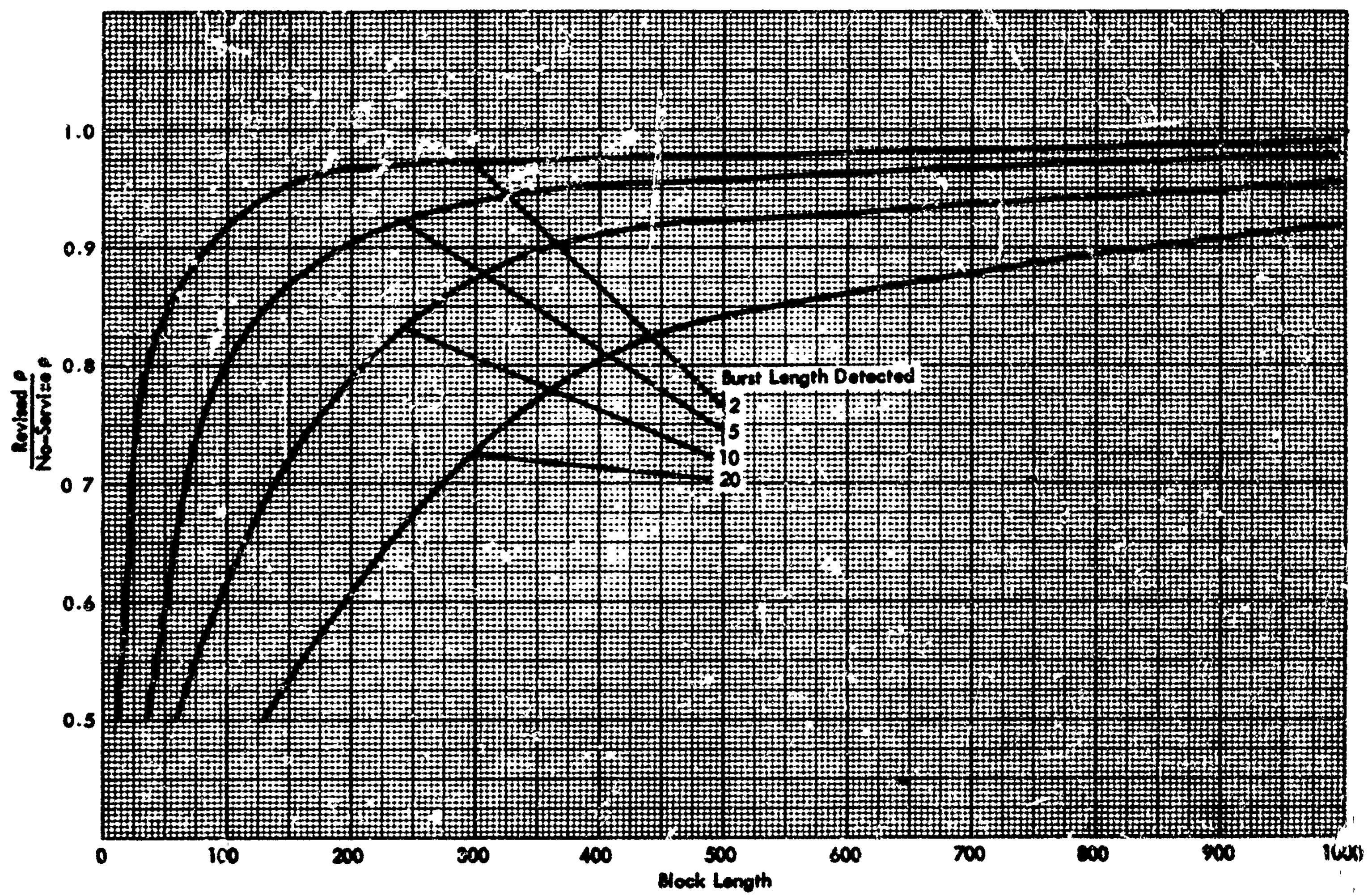


Figure 4-17. Decrease in Reliability with No-Service Dual-RQ, Burst Errors

introduction of errors into a message block. In practice, other control and service requirements in these blocks may preclude the use of No-Service Dual-RQ.

4.6.2 Non-Retransmission Logics with Feedback

4.6.2.1 General Discussion

The type of system to be considered here is shown in Figure 4-18. A buffer at transmitter and receiver is included to provide for convenient service to relatively constant rate sources and sinks of information. The data stream from the buffer is divided into blocks, and whatever redundant symbols are needed are added in the message formatter (for instance, symbols for synchronization, service information for the adaptive logic, and in some cases check symbols for error control coding). The transmitter section performs the necessary modulation of the symbols it receives, enabling their transmission to the distant terminal. At the terminal a receiver demodulates the signal and monitors the threshold variable, and the message processor removes the redundancy and recovers service messages generated by the other terminal.

Both the message formatter and transmitter receive signals from the adaptive decision block, as follows:

1. Observations of the threshold variable at the receiver (A) are used in the adaptive decision section (possibly along with error data from the processor (A)) to generate requests for a mode change in transmitting from B to A. These requests result in service messages originating at A's formatter.

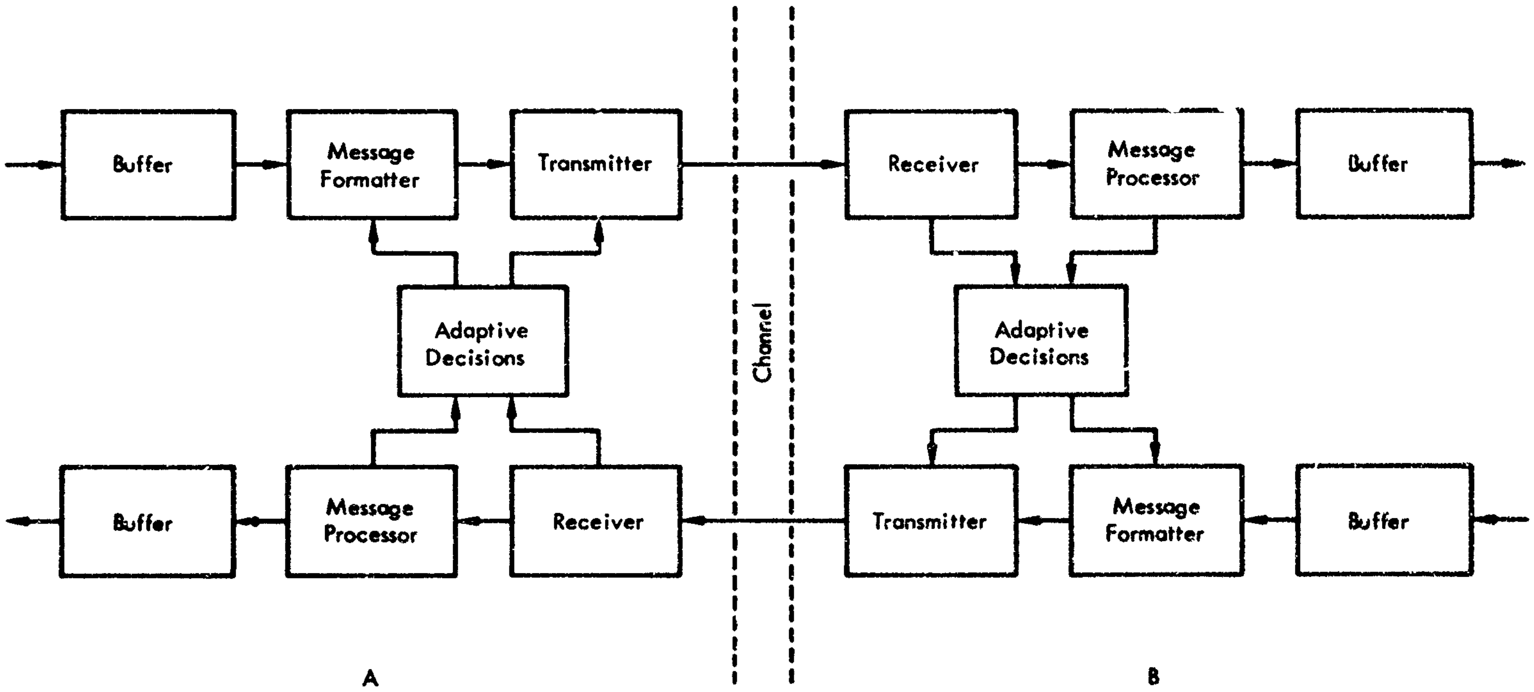


Figure 4-18. Adaptive System

2. When a service message is received at A, it is recovered at A's processor and routed by A's adaptive decision section to the formatter and transmitter. The transmitter is caused to change transmission mode (e.g., change speed of transmission), while the formatter makes changes in the redundancy, if needed, and possibly requests a change in the data rate from the buffer.

A basic assumption that will be made about all the systems considered here is that a clock is available at both terminals which provides absolute time information. Even when transmission speeds, etc., change, the basic time information is always available at both A and B. This may or may not imply phase coherence at the carrier frequency, but it does imply adequate timing for individual symbols and blocks of symbols.

In designing the adaptive control portion of the system, one of the most important considerations is the action taken by the receive terminal (B in this example), since we will assume that the transmit terminal (A) automatically obeys the commands it gets from B. In all the cases considered here, it will be assumed that the decisions determining the mode of the link will be determined at the receiver. This is logical in most simple cases, since the receiver has all the relevant information. Thus, only the decision need be transmitted back to the sending end, instead of detailed results of measurements.

It is possible that in some cases (not considered here) it would be advantageous for the decisions to be made at the transmitter. For instance, the transmitter is closest to the source of the information, and has knowledge of message priorities, anticipated traffic conditions, overloading of a message buffer, etc. In such cases it may be advantageous to transmit more information back to the sending end from the receiver, letting the decision be made at the transmitter site. In such a case, however, the information transfer rate is not being maximized, but more general criterion is employed, including the other factors mentioned.

Several alternative policies which can be followed at the receiver are listed.

1. The receiver can continue to receive in the old mode until enough time has elapsed for transmissions to change, then change receiving mode, assuming the transmitter has correctly interpreted the request. This is the most optimistic policy, and provides the minimum delay in the event no errors were made in the service messages. This can be accomplished either at the end of blocks of data, or after arbitrary transmission symbols, in which case the data may or may not be divided into blocks.

2. The receiver continues to receive in the old mode until there is positive indication that the transmitter has changed to the new mode. This has the disadvantage that if blocks are protected by an error-detection code, a change of state from good to bad will be accompanied by increased uncertainty of reception just when the new mode must be recognized.

3. The receiver could have two or more demodulation units, each constantly monitoring the channel and available to take over the job of receiving data as soon as transmissions are received in its mode. This may be optimum, in the sense that it allows for a random delay after a request to adapt before the actual transition is made. However, the scheme is wasteful of equipment, since all but one mode will need an idle receiver; and to aid in recognizing the mode change, division of the data stream into blocks with error detection coding would probably be necessary. Thus, a block with no error would be accepted, all others would be rejected, and if more than one block had no error detected (an event which would almost never occur), all could be rejected.

4. As a modffication of the first method above, the change could be requested in advance of the planned time for change, and several successive messages could be sent, each indicating the time for desired change. For instance, as soon as a threshold is crossed and it is decided that a change is needed, a request could be sent in the next block requesting a five-block delay before putting the change into effect. The next block sent from the receive station could request a four-block delay, etc. This scheme uses more time in the inefficient mode before changing, but has less chance of wasting data and time by errors in the control data. Obviously, this scheme would require relatively long periods between transitions to be efficient. This scheme seems promising, and will be elaborated in subsection 4.6.2.2.

In a realistic evaluation, however, it is important to consider the frequency with which mode changes are made. To analyze the problem in this light, the second-order statistics of the threshold variable must be considered. For instance, the number of times per second that a given threshold is crossed in the positive-going direction is important in determining thresholds, as well as the mean duration of intervals for which $\theta_{i+1} > \theta > \theta_i$ for various fixed thresholds θ_i. (See subsection 3.4.)

4.6.2.2 A Controlled Delay Adaptive Logic

A full-duplex scheme is elaborated here which changes modes according to decisions made at the receivers, after a delay which is known at the receiving terminal. The basic concepts of this adaptive scheme are:

1. Block and clock synchronization at both terminals are assumed perfect.
2. All blocks are of the same duration (in time). Symbols may be of variable duration and alphabet size.
3. There is no attempt at retransmission. If error coding is used, erroneous blocks can be discarded.
4. A threshold variable will be observed at the receiver, and the events of crossing thresholds will be noted.

Initially the logical control of a three-mode adaptive system will be examined. Although a two-mode adaptive scheme would be simpler, such a scheme is qualitatively different in character from a multi-mode system, and results could not be generalized. The three-level system encounters not only the problems of communicating a request for a mode change, but also the problem of identifying <u>which</u> mode change. In changing modes, a system with ordered modes can have a policy of changing by only one mode at a time, or of allowing "jumps" to new modes in order to adapt more quickly to rapid changes in the channel. Another reason for starting with a three-mode system is the intuitive belief that this will allow a central region or zone for the threshold variable, in which more or less average performance can be obtained for a large fraction of the total time. Then, deviations above or below this zone would result in adapting the mode, either to salvage some performance in a period of poor S/N ratio, or to capitalize on a period of very good S/N ratio by sending data more rapidly.

On crossing the (i-j) threshold, indicating that the channel has changed from state i to state j, a service message will be inserted in the earliest possible block sent back to the other terminal (the terminal transmitting over the changing channel). The service message will demand a change from mode i to mode j after k blocks. Succeeding blocks sent in that direction will include service messages demanding a change from mode i to mode j after k - 1, k - 2, ..., etc., blocks. The transmitter, on receiving the block with service message requesting mode j after zero blocks, will commence transmission in mode j in the next complete block it transmits. An example of successive service messages might be

(3,4), (3,3), (3,2), (3,1), (3,0), (3,0), (3,0), ...

Here, k is 4, the channel has changed from some unknown state i to state 3, and the request is to change to mode 3 after 4, 3, 2, 1 and 0 blocks. Note that it costs nothing to continue the service message (3,0) until the next state change, thus enabling a recovery after an error.

Because of the assumed synchronization, the terminal demanding a mode change can switch its receive mode without delay the instant transmissions are expected in the new mode (if there is no changeover time for the equipment).

In case of error in interpreting a mode change request on the part of the transmitter terminal's receiver, the receiver will continue in the new mode in anticipation of an eventual recovery at the transmitter. Similarly, in case of a false-alarm at the transmit terminal, where a mode change is thought to be requested but in fact has not been, the receiver takes no action, but waits for the transmitter to uncover the mistake.

In the event another threshold crossing occurs before the full sequence of service messages is sent, the sequence will be cut short and a new sequence begun, requesting a change to the mode corresponding to the new state of the channel. This would lead to such a service message sequence as:

(1,0), (1,0), (1,0), (2,5), (2,4), (2,3), (2,2), (3,5), (3,4), (3,3), (3,2), (3,1), (3,0), (3,0), ...

At the transmitting terminal, whenever a command to change is received but an error is known to exist in the block bearing the command to change, or whenever an error is known to exist and no change request has been received, decisions must be made about what mode change to make and when to make the change. The possible strategies available at the transmit terminal are outlined as follows:

If error detection coding is not available, a block can be received with an allowable service message which is not in proper sequence, indicating an error, though not necessarily in that block. For instance, if $k = 4$, and the mode has been 1, if a service message sequence such as (1,0), (1,0), (2,3), (2,2) (2,1), (2,0), (2,0), ... occurs, the transmitter can reasonably infer that the last (1,0) service message should have been (2,4) instead of (1,0).

A similar situation would be a sequence of service messages such as (1,0), (1,0), (2,4), (1,0), (1,0), ... Here again, the inference of an error is reasonable since the entire sequence (2,4), (2,3), (2,2), (2,1), (2,0) was not received, and the transmitter would make no change. Of course, a completely unallowable message such as (7,5) could always be recognized as an error. If $k = 0$, then a sequence (1,0), (1,0), (3,0), (3,0) would necessarily be accepted, unless mode "jumps" were prohibited. In that case (1,0), (1,0), (2,0), (3,0), (3,0) would be a proper sequence; or if a minimum time in each mode is required, e.g., three blocks, a proper sequence would be (1,0), (1,0), (2,1), (2,0), (2,0), (2,0), (3,0), (3,0), ...

In a period of numerous errors, erroneous service messages can be expected very often, perhaps 1/2 of the total number of service messages being in error. In the absence of error detection coding, a value of k greater than zero seems to be the only available protection against indiscriminate erroneous mode changes which are unpredictable at the receiver terminals. Although there is some evidence of correlation between the fading of two channels operating in opposite directions over the same medium [64], it is not sufficient for a detailed system design. Only if large correlation were established would a "fail-safe" policy of reverting to a lower performance mode on receipt of an error be justifiable. This is because errors can occur (although they are more unlikely) even in the best possible channel state, and the fail-safe logic involves participation by both terminals, each basing its decisions and actions on estimates of the action at the other terminal. It is believed that only for a serial retransmission system, where erroneous blocks are retransmitted immediately and in their proper sequential order, is such complicated logic justified.

With detection of errors by error detection coding techniques, the confidence in a service message decoded at the transmitter is excellent if no error is detected, since efficient codes can yield a probability of an undetected error on the order of 10^{-20} or less. If a block error is detected, which in a bad fade could occur with probability greater than 0.5, the transmitter terminal may still accept the data corresponding to the service message. Here, the same devices as before can be used to recognize obvious errors (for instance, prohibiting jumps from a mode to a non-adjacent mode, or allowing only certain sequences of service messages will enable some errors in the service message to be detected within an erroneous block). The service messages in an erroneous block which are not obviously erroneous could then be assumed correct, protected by an error correction code, ignored (thus continuing transmissions in the old mode until a correct block is received), or examined for two or more successive blocks, so that the most likely sequence of service messages could be accepted, but adding delay in the transition time.

This system can recover from false mode changes and delayed mode changes. In all cases the receive terminal will be passive, not attempting any correction except to continue its service message: (i,0), (i,0), (i,0), ..., requesting the i^{th} mode, as long as its sensors indicate that the channel over which it is receiving transmissions remains in the i^{th} state.

At the transmit terminal (A), knowledge that (B) is not confusing the situation by trying to second-guess A's motives and actions is of inestimable value. Thus, a recovery is certain as soon as a block is received by A which has no error (as indicated by an error-detecting code). The expected number of blocks required to obtain an error-free block is (1/p), where p is the probability of an error-free block on each trial (assuming independence, etc.). Thus, even with a block error probability of 0.9, an error-free block would be expected to occur after an average of nine erroneous blocks (i.e., in the tenth block).

4.6.3 Adaptive Systems Without Feedback

The following is a description of a specific type of forward error correction system applying variable utilization of fixed redundancy. The state of the channel is monitored by examining the coding structure of received messages. The decoding procedure is adapted to perform either burst error correction or random error correction.

Both the channel monitoring and the burst error correction use a technique, statistical burst correction, recently developed by IBM on an internally funded project. This technique differs from other coding techniques in that it uses far fewer redundancy bits for the correction of burst errors of a specified length. This means that by employing the new technique it is possible to obtain higher throughput and/or greater frequency of burst error correction.

A distinguishing characteristic of this technique is that, although it corrects almost all of the burst errors of the length for which it was designed, it does not correct all the bursts of that length.

Let P(x) denote the code polynomial for a code of block length n. Let r denote the degree of P(x). If r is chosen so that

$$r > b + \log_2 n, \tag{4-23}$$

then the code generated by P(x) can be used to correct bursts of length less than or equal to b bits. Note that the choice of P(x) is unimportant; only the degree of P(x) is critical. A standard polynomial encoder is used in implementing this technique. To implement the burst correction, construct the syndrome or check word associated with each block received in the decoder so that it represents the modulus (with respect to P(x)) of the error pattern in the received block. Write the syndrome S(x) in terms of those powers of x corresponding to the first r bits of the received block, that is

$$S(x) = \sum_{i=1}^{r} s_{r-i}\, x^{n-i} \tag{4-24}$$

Note that if none of the errors in the block have occurred outside the first r bits in the block, then the check word will contain the exact error pattern. In particular, if none of the errors have occurred outside the first b bits, then the coefficients of x^{n-i}, $i=r-b+1, \ldots, r$ in S(x) will be zero. If errors have occurred outside the first b bits, but none have occurred outside the first r bits, then some 1's will appear in the coefficients of

x^{n-i}, $i=r-b+1, \ldots, r$, in $S(x)$. If errors occur outside the first r bits in the block, then there is some probability Q (dependent upon the polynomial and on the number of errors outside the r bits) that each single coefficient of x^{n-i}, $i=r-b+1, \ldots, r$, in $S(x)$ will be zero. If the probability for each of the coefficients is independent, then the probability that all of the stages from 1 to $r-b$ will be zero simultaneously is equal to Q^{r-b}. Thus, if the condition of all zeros in the coefficients of x^{n-i}, $i=r-b+1, \ldots, r$, in $S(x)$ is used to indicate a burst of errors in the first b bits of the block with the pattern of 1's in $S(x)$ representing the error pattern, then a false correction will be made with probability Q^{r-b}. Since Q is less than unity, Q^{r-b} can be made arbitrarily small by making $r-b$ large.

If the condition for the existence of a burst is not met, then the syndrome can be rewritten in terms of $x^{r-2}, \ldots, x^{n-r-1}$, and the same condition checked to determine if a burst of errors has occurred in bits 2 to $b+1$ of the block. Each succeeding shift of the syndrome will shift the span of bits over which the burst is detected by one bit position. The probability of falsely detecting, at some shift through the block, the condition which indicates a burst of length less than or equal to b bits is then $(n-b)Q^{r-b}$. For large values of b, this technique provides the potential for extremely efficient and effective burst error correction.

Recall that the statistical burst correction technique does not require any particular type of polynomial, but only one having degree somewhat greater than the correctable burst length. In particular, a polynomial that has random error correcting capabilities can be used. With such a polynomial, a system incorporating the variable use of redundancy is easily constructed. The statistical burst correction technique is used to determine whether or not a correctable burst has occurred. If a correctable burst has occurred, it is corrected. If a correctable burst has not occurred, the syndrome can then be applied to random error correction.

The selection of a very high degree polynomial with random error correction properties that can reasonably be implemented presents a problem. A solution is obtained by interleaving a large number of low degree polynomials which have random error correcting properties. If $P_1(x)$ is a polynomial of degree r_1, then $P(x) = P_1(x^t)$ is a polynomial of degree $r = tr_1$ which is equivalent to interleaving t subwords, each encoded with $P_1(x)$. If $P_1(x)$ has good random error correction capability in a block of n_1 bits, then $P(x)$ has good random error and multiple short-burst error correction capability over a block of length $n = t\,n_1$. Such a polynomial $P(x)$ can also be used for statistical burst correction with adaptive decoding.

The use of such an interleaving polynomial $P(x)$ does present some special problems for statistical burst correction and adaptive decoding. If the correctable burst length b is not chosen so that $r-b \leq t$, then errors in the guard space (i. e. the remainder of the block outside the burst) in certain of the interleaved words can occur without causing 1's to appear in the last $r-b$ bits of the check word while the burst pattern is in the places corresponding to the first b bits of the check word. This can result in false correction. Furthermore, a small number of random errors or a short burst of errors may also result in a false indication that a burst has occurred; that is, it may result in the last r-b bits of the check pattern being all - 0's, while the first b bits do not display the actual error pattern. Thus, some additional criteria must be included in the adaptive decision process to determine whether burst error correction or random error correction is appropriate.

One set of additional criteria which is sufficient for determining that a burst has occurred when the last r-b bits of the check word are zero is the following:

1. $r-b \geq t$
2. The number of syndromes of interleaved subwords which correspond to error conditions other than single errors is greater than some specified threshold.

The choice of the appropriate threshold for condition 2 presents some interesting analytical considerations. The number of random errors which can be corrected in each subword is critical to the choice of the threshold. Suppose the number of random errors that can be corrected by using $P_1(x)$ is e. Then the threshold should be chosen high enough so that it presents a good indication that a large number of errors have occurred (as in a burst of length greater than et). However it should not be so high that the threshold will not be exceeded when many of the subwords have more than e errors. When t is small a count of the subwords with any errors at all may also be helpful.

4.7 IMPLEMENTATION OF ADAPTIVE TECHNIQUES

4.7.1 Non-Error-Control Techniques

In some of the adaptive systems being investigated, the data rate is varied by changing either the symbol length or the size of the symbol alphabet. An understanding of the operation of modems is necessary at this point in order to determine whether or not commercially available modems can accommodate changes in these parameters and, if not, what modifications are necessary in order to implement these adaptive techniques. The question here is: While these are good systems theoretically, are they feasible with respect to implementation and cost?

4.7.1.1 Modem Requirements

The ideal modem for this application must be able to transmit and receive data using different symbol lengths or symbol alphabets and to automatically switch among these different modes upon command. Further, the modem must perforce maintain exact synchronization during these transitions and not require any time for resynchronization. A survey of commercially available modems was made in order to determine whether or not any available modems could satisfy these requirements. Some of the modems investigated and their operating characteristics are listed in Table 4-1. It was found that, while many modems offered different operating speeds, none could be instantaneously varied upon command. No modem was found that offered a choice in alphabet

Table 4-1

Modems and Operating Characteristics

Modem	Land Line Tropo. Micro-wave	HF	Maximum Bit Rates (b/s)	Full Duplex (F). Half Duplex (H)	Number of Subchannels within 3kc band	Type of Modulation	Diversity Reception Available	Commercial or Experimental	External Clock Option	Data I/O - Serial and or Parallel
Tele Signal Model 2036	X	X	1440	F	16	FSK		C		P
Tele Signal Model 2062	X	X	1800	F	24	FSK		C		P
Tele Signal Model 2063	X	X	1440	F	16	FSK	X	C		P
Tele Signal Model 60/1		X	1800	F	16	FSK	X	C		P
Tele Signal Model 60/25A		X	1800	F	16	FSK	X	C		P
Rixon Sebit - 24B	X		600, 1200, 2400, 500, 2500	F, H	1	AM + PSK		C	X	S
Bell Data Set 201A	X		2000	F, H	1	Differentially Coherent PSK		C	X	S
Bell Data Set 201B	X		2400	F, H	1	Differentially Coherent PSK		C	X	S
Bell Data Set 301B	X		40.8K	F, H	*	Differentially Coherent PSK		C	X	
Collins TE-216A-20		X	4500	F	60	PSK	X	C		S, P
Collins TE-216A	X	X	600, 1200, 2400, 3600, 4800, 6200	H	60	PSK	X	C		S, P
Collins TE-216A-4	X	X	600, 1200 2400, 3600	F, H	**	AM + PSK	X	C		S, P
Milgo Series 4018-30, 31	X		2400			FSK		C	X	S
Rixon DD-1C03		X	1200		16	FSK		C		S
Gen. Dyn. SC-302		X	3000		40	PSK		E		S
Gen. Atronics Kathryn		X	3000		40	Differentially Coherent PSK		F		P
Lenkurt 27A		X	2400		16	Noncoherent FSK	X	C		P

*Wideband TELPAK A channel
**2, 3, 4, 6, 8, 9, 12 subchannels

size. In fact for any given medium, most manufacturers used almost identical alphabet sizes, and, in most cases, the maximum number of bits per symbol was two. In some of the extremely high speed modems, three bits per symbol were used.

Since it appears unlikely that commercial modems will satisfy the requirements for our particular adaptive systems, we will investigate the problems and costs involved in designing such a modem or modifying existing modems.

4.7.1.2 Variable Symbol Duration

Typically, modems operate in the following manner. Data, in a serial bit stream, is obtained from a source and gated into a modulator as shown in Figure 4-19. The output of the modulator, an analog signal, is then passed through a 3-kc band pass filter, thus limiting the spectrum to the allotted voice channel bandwidth. (Although channels with larger bandwidths are used, this discussion will be limited to a 3-kc nominal voice channel.) The output of the filter is then fed to a transmitter. At the receiver, the incoming signal is passed through a filter, demodulated, and gated into the data sink. The bandwidth of the filter at the receiver is a function of the data rate and the modulation used. Now, if the data rate were to be suddenly cut in half by doubling the symbol duration, the bandwidth of the data signal would be halved. Thus, the filter at the receiver must be changed to match the reduction in bandwidth in order to optimize the noise rejection. If the filter is not changed, then the minimum required channel signal-to-noise ratio will be larger than if the filter is changed. This is due to the fact that noise outside the bandwidth

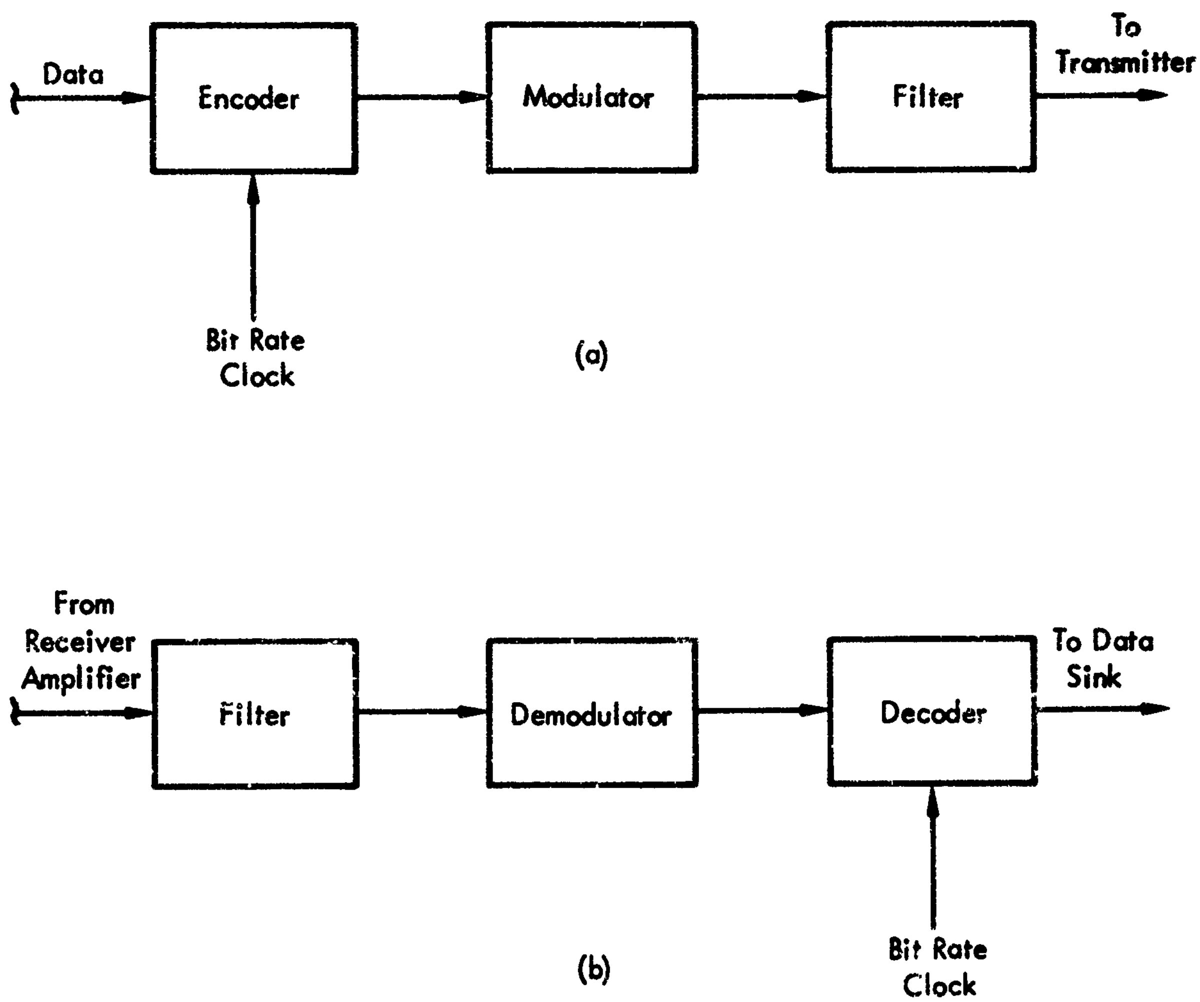


Figure 4-19. Modem at (a) Transmitter, (b) Receiver

of the signal will pass into the detection circuits. It seems reasonable to assume that the data rates selected for each mode of the adaptive system will be equal to $(75)2^K$. Thus, if the filters at the receiver are not changed, a loss will be suffered each time the bit rate is halved. Thus, it appears that a modem designed to handle different data rates should include filters matched to the spectrum of the signal at each data rate.

For single channel modems, this is not too unreasonable. The number of different data rates will probably be around two or three, and therefore the extra hardware involved will be small compared to the design and modification costs. However, in modems designed for use in the HF region, the increase in cost becomes considerable due to the frequency division multiplexing schemes usually used. In HF modems the input high speed serial bit stream is usually converted into n parallel slow-speed bit streams so that the information symbol duration will be sufficiently long to counteract the intersymbol interference due to multipath smearing. Each of these parallel bit streams is then fed to its own modulator, and at the receiver each has its own receiving filter. Thus, filters in each of the parallel subchannels must be switched, and the cost of the additional filters will be n times as much as for the single channel. Note that if the bit rate of each of the n parallel channels is $(75)2^K$, then the overall bit rate becomes $n(75)2^K$. This allows a little more flexibility in selecting overall channel bit rates since rates that are multiples of $(75)2^K$ can now be obtained.

4.7.1.3 Variable Alphabet Size

In the adaptive system being considered here, the data rate will also be varied by changing the symbol alphabet, i.e., number of bits per symbol, while keeping the symbol duration constant. Again, this change will be accomplished instantaneously at both the transmitter and receiver.

Modems developed for wire lines and telephone circuits may also be used in microwave, tropospheric scatter, cable and VHF and UHF line-of-sight radio circuits. In these modems it appears, judging from the survey of commercial equipment available, that the modulation schemes very seldom employ symbol alphabets greater than two bits per symbol. For example, in phase modulation, the most popular modulation used, adjacent pairs of bits are encoded into one of four phases. Modems employing an eight-phase (three bits per symbol) modulation scheme have not been found. Some experimental modems use a combination of phase and amplitude modulation to achieve a symbol alphabet conveying three bits per symbol.

Thus, it appears that in an adaptive communications system that depends upon changes in symbol alphabet size to vary the bit rate, the state of the art limits the choice to either one, two, or three bits per symbol.

A system using simply two alphabet sizes is feasible and does not appear too difficult to implement.

Techniques used in both frequency shift keying (FSK) and in phase shift keying (PSK) are discussed in the following paragraphs.

4.7.1.3.1 Four-Level FSK

A typical four-level FSK system is shown in Figure 4-20. The data in a serial bit stream is gated into the modem two bits at a time and fed to a digital-to-analog converter. For purposes of discussion, the first bit in the bit pair is called the "even" bit and the second the "odd" bit. The resulting voltage level is converted to one of four tones by a voltage controlled oscillator (VCO), passed through a band pass filter and sent to the transmitter. Note that each tone represents two bits of information. At the receiver the signal, after amplification and filtering, is passed through a limiter and a discriminator. The output of the discriminator is one of four voltage levels. This level is then converted to a two-bit digital number in an analog-to-digital converter, and the binary digits thus obtained are gated to the data sink. The transfer characteristics for the VCO and discriminator are shown in Figure 4-20. Table 4-2 shows the A-D output levels and tones produced for each of the four possible bit pairs.

Table 4-2. Input Bit Coding for a Four-Level FSK System

Input Bit Position		Output of D - A	VCO Output Tones
Even	Odd		
0	0	V_1	f_1
0	1	V_2	f_2
1	1	V_3	f_3
1	0	V_4	f_4

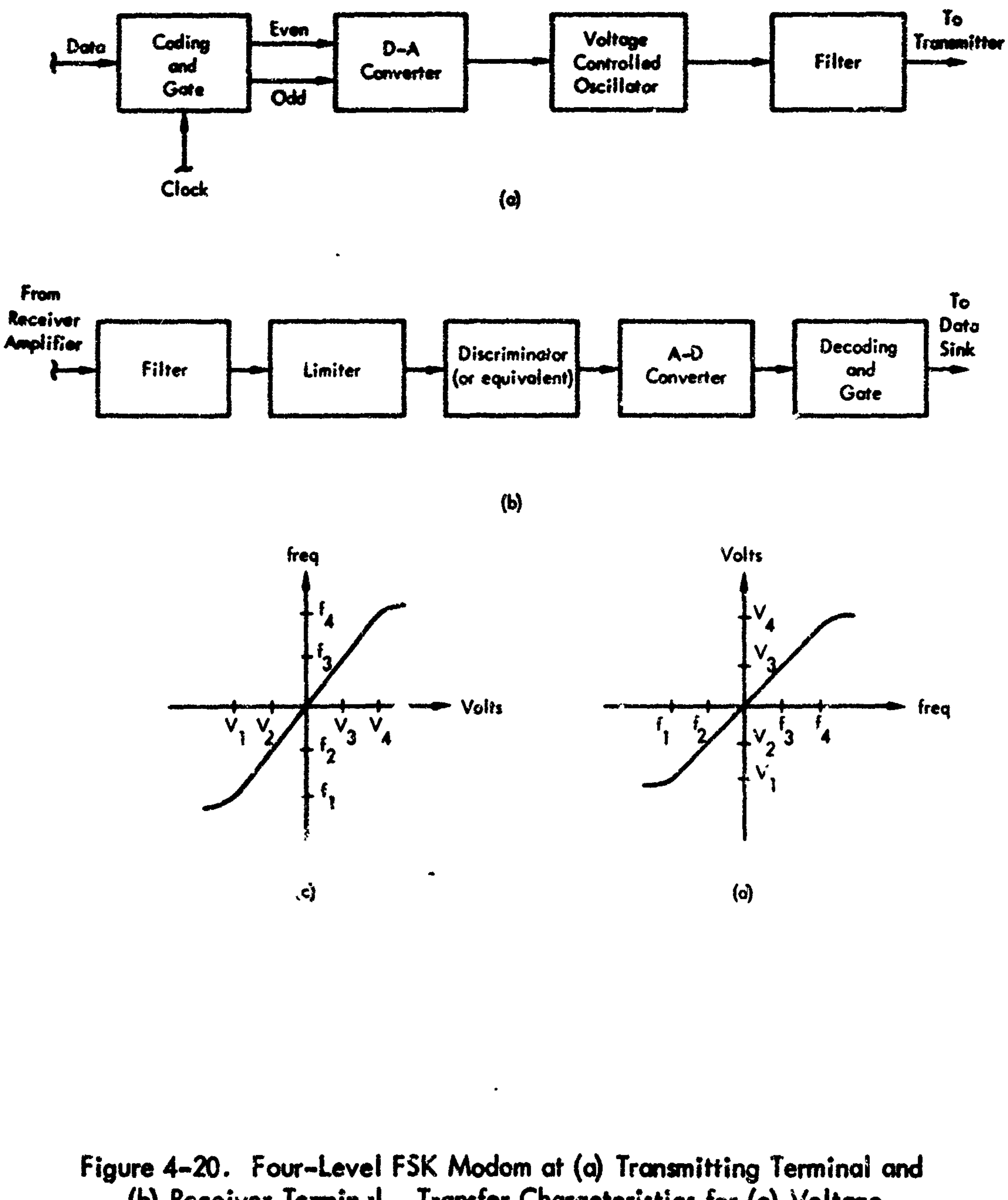


Figure 4-20. Four-Level FSK Modom at (a) Transmitting Terminal and (b) Receiver Terminal. Transfer Characteristics for (c) Voltage Controlled Oscillator (d) Discriminator.

A cyclic Gray code is usually used because it produces fewer bit errors than a natural binary code.

4.7.13.2 Two-Level FSK

For adaptive systems, the four-level system described in the preceding paragraph can easily be converted to a two-level system. In Table 4-2, if the odd input bits are set to zero, and only the even input bits are allowed to change with the data, then the output of the D-A converter will vary between the two voltage extremes V_1 and V_4, and the VCO output would be either of the widely separated tones f_1 or f_4 (see Figure 4-20c). The input data stream would naturally have to be cut in half and would essentially be fed directly to the even bit input of the D-A converter. The odd bit input would be tied to a logical zero. The same effect can be accomplished at the data source merely by inserting the data into even bit positions and zero's into the odd positions.

The data sink can retrieve the data in either case by simply rejecting all bits in the odd positions. Note that the symbol length has remained constant.

From Table 4-2 we can see that the reliability of the two-level system is greater than the four-level system. In the four-level system, if a transmitted tone f_1 is received as tone f_2, an error has occurred. In the two-level system a transmitted tone f_1 would have to be received as tone f_3 in order for an error to occur.

Note that in this system the clock does not change. We can switch back and forth between four-level FSK and two-level FSK without any change in the clock rate, thus eliminating any synchronization problems.

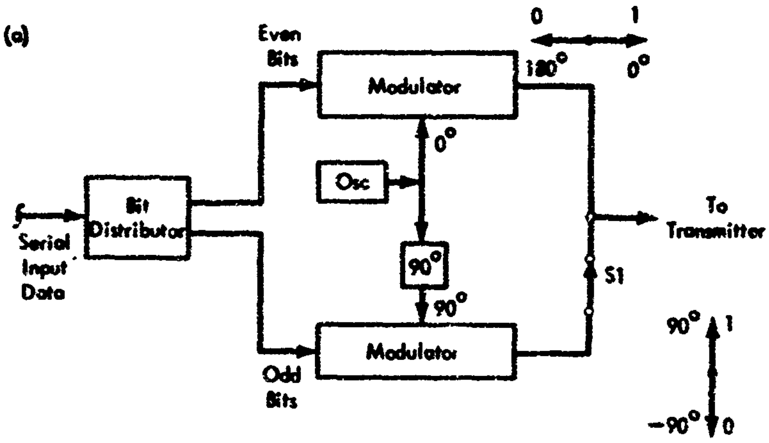
The cost of implementing this system would be very little; all that is required is some additional logic in the transmitting and receiving terminals.

For HF modems, a serial-to-n-parallel conversion is performed upon the input bit stream as previously described. In this case, the discussion above would describe the operation of one of the n subchannels, the only difference being that now the serial-to-parallel conversion must be performed on pairs of bits rather than on single bits. This is necessary to ensure that each of the n parallel bit streams can be treated as a separate channel transmitting bit pairs. The cost of a special serial-to-parallel converter to accomplish this is small and can easily be absorbed in the logic design of the transmitter and receiver adaptive terminals.

4.7.1.3.3 Four-Level PSK

Figure 4-21a shows a common method of generating a four-phase signal. The input serial binary data is separated into even and odd bits and routed in parallel to the two PSK modulators. The output of the upper modulator is a sine wave whose phase is either 0° or 180° with respect to a reference, depending on whether the input is a "one" or a "zero," respectively. Similarly, the output of the bottom modulator is either +90° or -90° out of phase with a reference, depending on whether the input is a "one" or "zero." The output of these two modulators is added and the resulting wave is filtered and applied to the transmitting circuits. Figure 4-21b is the "truth" table for the resultant output phase as a function of the input pairs of bits. Again, a cyclic Gray code is used.

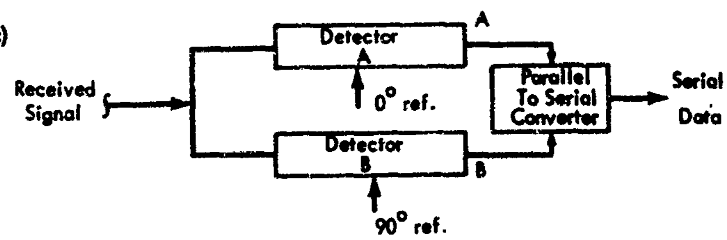
(a)



(b)

Input Bit Position		Resultant Output Phase	
Even	Odd		
0	0	-135°	ϕ_1
0	1	+135°	ϕ_2
1	1	+ 45°	ϕ_3
1	0	- 45°	ϕ_4

(c)



(d)

Input Phase	Binary Code A	B	A 0° Detection	B 90° Detection
+ 45°	1	1	+	+
- 45°	1	0	+	-
+135°	0	1	-	+
-135°	0	0	-	-

Figure 4-21. (a) Generation of Four-Phase Signal (b) Truth Table Describing Resultant Phase for Each Pair of Input Bits (c) System for Detecting Four-Phase or Modified Two-Phase Signal (d) Truth Table at the Detector.

A system for detecting four-phase signals is shown in Figure 4-21c. The incoming signal is compared to a fixed reference phase of 0° or 90° and converted to a two-bit binary number. Figure 4-21d shows the relationship between the input phases, their digital assignments and the polarity of the detector outputs. Note that the output of detector A will be positive (resulting in a one in bit A of the binary code) whenever the input signal has a component that lies in either the first or the fourth quadrant, such as ±45°. The output at A will be negative (resulting in a zero in bit A of the binary code) when the input signal has components that lie in the second or third quadrant such as ±135°. The parallel to serial converter in Figure 4-21c converts the parallel digital outputs of detectors A and B to serial form as shown in Figure 4-21d.

4.7.1.3.4 Two-Level PSK

The four-level PSK system can be easily converted to a two-level system for use in adaptive systems by making a minor modification to the transmitting modem. Consider the table in Figure 4-21b. For a two-level system, the logical choice of phases to represent the binary data would be any two phases separated by 180°. Thus, either of the pairs ϕ_1 and ϕ_3 or ϕ_2 and ϕ_4 could be chosen.

Suppose for the moment the pair of phases ϕ_1 and ϕ_3 were chosen to represent a binary zero and one respectively. If either of these phases were received as ϕ_2, it would be impossible to tell whether ϕ_1 or ϕ_3 had been transmitted. Thus, no additional protection against random noise has been obtained by using only two phases of a four-phase system.

A simple modification to the modulator and demodulator can be made to provide this additional protection against noise. If the switch labeled S1 in Figure 4-21a is opened, the phase of the sine wave sent to the transmitter will be either 0° or 180° with respect to the local reference, depending upon whether the input data is a one or a zero. The receiver, shown in Figure 4-21c, need not be modified at all. So long as the input wave has a component lying in the first or fourth quadrant, the bit in position A of the output binary code will be a "one." It will be a zero if the incoming wave has a component lying in the second or fourth quadrant. Thus, since only waveforms of 0° and 180° phase will be transmitted, the output of detector A will either be positive or negative, indicating that the data bit in position A is either a "one" or a "zero." If the input phase is corrupted by noise, the output of the detector at B will either be positive or negative, depending upon whether the phase of the signal component orthogonal to the 0° reference is +90° or -90°. Thus, the bit in position B will take on values of zero or one at random. It is clear that this two-phase system enjoys a greater immunity to noise than a four-phase system.

Since the modulator receiving the odd bits in Figure 4-21a is switched out of the circuit during two-phase PSK, the input data rate can be reduced by one half by inserting the data in the even positions and zeros in the odd positions of the input bit pairs. A data stream of 101, for example, would be coded by the source as 100010 prior to entering the modem. In this way, the bit distributor in the transmitter modem would remain the same. The receiving terminal can easily restore the data to its original form by simply rejecting all odd bits.

The switch S1 in Figure 4-21a would be replaced by an electronic gate controlled by the transmitter terminal. The gate would be either conditioned or de-conditioned, depending on whether the system is operating in the four-phase or two-phase mode.

In this system, the clock need not be modified, thus eliminating any synchronization problems. Note also that the symbol length has remained constant. The cost of implementing this system may be considerable, since a change in the modem is required. The costs of the required special data handling can easily be absorbed in the logic design of the transmitting and receiving terminals.

The comments that were made in subsection 4.7.1.3.2 with respect to HF modems apply here also. A serial to parallel conversion must be performed on bit pairs rather than on each individual bit.

A four-phase system has a three-db less margin to noise than a two phase system for a given bandwidth and symbol rate.

4.7.1.4 Synchronization

For efficient detection, a reference wave is required at the receiver that is in phase with the reference wave used at the transmitter. This is necessary to provide reliable mid-bit data sampling.

In adaptive systems, a further requirement is made of the receiver—that is, the receiver must not lose synchronization when the data rate is changed. The survey of available modems uncovered no modem that could meet this requirement. In almost all cases, a start-up time was required to enable the receiver clock to lock in on the incoming data.

In Figure 4-22, a block diagram of a phase corrected clock is shown. This is a modified version of a clock that has been built by IBM and proved successful. This modified version could be used in an adaptive data communications system involving three data rates $(75)2^{K}$, $(75)2^{K-1}$ and $(75)2^{K-2}$.

A high stability oscillator provides the necessary pulse source. This oscillator is N times the highest bit rate where N is a large integer (typically several hundred). The data from the demodulator is fed into the phase error detection logic. This logic compares the phase difference between the input digital data and the locally generated clock. When a phase difference is detected, a signal to either retard or advance the phase of the local clock is given to the phase corrector logic. This logic modifies the pulse train from the local oscillator to the divide-by-N counter by either inserting or inhibiting a pulse, depending on whether the phase of the local clock is to be advanced or retarded. The clock selection and gating logic instantaneously change the frequency of the local clock upon orders from the receiving terminal. This system, once synchronized, will not lose synchronization when the data rate is changed.

The cost of implementing this system is small. It is possible to design a modem which can operate at several different speeds and can be clocked externally. The system described here could be used to supply this clock.

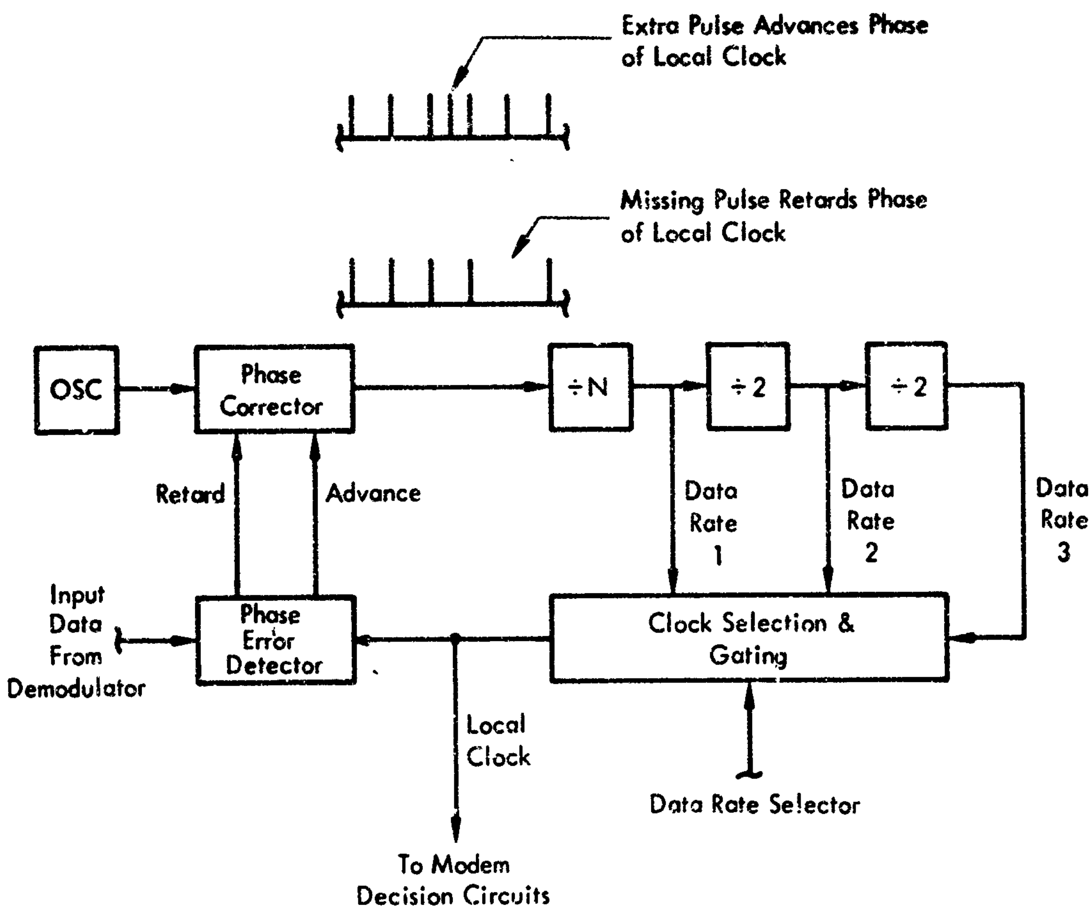


Figure 4-22. Block Diagram of a Phase-Corrected Clock for Use in a Three-Speed Adaptive Data Communication System.

4.7.1.5 Conclusions

The feasibility of implementing certain adaptive communications systems has been established. A survey of data modems commercially available was taken, and it was concluded that modifications would be necessary in some cases to implement any of the adaptive systems under consideration here.

For variable symbol length systems, it was found that to maintain good noise rejection at the different speeds different filters had to be switched in and out. Aside from the cost of the additional filters, the cost of modifying an existing modem to accomplish this would be considerable. Also, the cost gets larger when HF modems are considered.

In another system, the bit rate was varied by changing the alphabet size from four-level to two-level while maintaining a constant symbol length. Both FSK and PSK modulation schemes were investigated. It was found that both these modulation systems had the advantage of being easy to implement with very little modification to existing modems. The variable alphabet systems are particularly attractive, since the clock rate remained constant, eliminating any clock synchronization problems.

4.7.2 Error Control Techniques

This subsection considers the problems encountered in implementing the three basic adaptive error control techniques:

1. Variable block length, fixed redundancy
2. Fixed block length, variable redundancy, and
3. Variable use of fixed redundancy with a fixed block length.

These systems will be discussed with respect to feasibility of encoding, decoding, and maintaining synchronization. For the first two systems, it will be assumed that the control logic needed to maintain both encoder and decoder in the same mode is provided and is the type described in subsection 4.6.1 or 4.6.2.

4.7.2.1 Variable Block Length, Fixed Redundancy

Designing a variable block length encoder and decoder by modifying a system designed to handle the largest desired block length is relatively simple. In a standard encoder, the redundancy is calculated over a number of information bits determined by sensing a value in a counter. To decrease this number it is sufficient to incorporate controls so that the encoder will sense a specified smaller number in the counter when a mode change is made and return to sensing the original value when the mode is changed back again. The problem in the decoder is only slightly more complex. Again, it is necessary to change the value sensed in a counter to indicate the length of the block and thereby also indicate when the check word or syndrome has been constructed. In all systems, except for a few error-detection-only systems, it is also necessary to store the information bits in the decoder until the checkword has been constructed and then retrieve the bits in the proper order. This requires modifying the controls on the storage element used. If the storage is a core memory, this is trivial. It is necessary to change only the sensing of a maximum address before resetting the address register to zero.

Synchronization for this technique presents some difficulties unless the timing of the mode changes at both the encoder and decoder can be precisely coordinated. One method for obtaining initial synchronization is to use the encoded structure of the message [12]. The advantages of this technique are that it does not take additional channel space away from information bits and that whenever it is not received correctly the first time, it can be detected on the second, third, or subsequent transmissions. However, a resynchronization requirement when going to a shorter block length is undesirable: the reason for going to a shorter block length is that more errors are occurring in the channel. Therefore, resynchronization is more difficult. The most desirable technique still appears to be use of the coded structure of the message, eliminating as much of the problem as possible by coordinating the time of mode change in the encoder and decoder.

4.7.2.2 Fixed Block Length, Variable Redundancy

A fixed block length system will be simpler to implement than a variable block length system. None of its problems will be of the same order of difficulty as the synchronization problem described in the previous section. In fact, in a fixed block length system there is no synchronization problem except for a "start-up" procedure which is required even for a nonadaptive system.

The encoder and decoder for a fixed block length, variable redundancy system must handle a different number of information bits in each mode. Thus, they include the same problems with the same solutions described in subsection 4.7.2.1. In addition, the registers for constructing the redundancy in the encoder and the checkword in the decoder must be modified for each mode change. Since the capability to implement the mode with the most redundancy must be included, this becomes a simple change when using polynomial coding. It is merely a matter of changing feedback paths in a shift register with "exclusive-or" circuits between the stages. If the mode change is not perfectly coordinated at the encoder and decoder, the decoder will get a checkword indicating that errors have occurred. The only additional circuits needed are the added "exclusive-or" circuits and the gating to control which feedback path should be used.

4.7.2.3 Variable Use of Redundancy

A system incorporating variable use of fixed redundancy in a fixed block length is not faced with the synchronization problems described in subsection 4.7.2.1 or with the variable storage and feedback problems of encoding and decoding. It must, however, include the correction circuitry for two types of correction, a complexity comparable to the other two systems. This system, however, is not faced with the problems of coordinating mode changes between the encoder and decoder since the encoder operation remains constant.

A forward error control system of the type described in subsection 4.6.3 has been designed and fabricated (thus answering the question of feasibility) and is now in operation at the IBM Engineering Laboratory. The first model of this technique utilizes a block length of 3200 bits, equally divided between data and redundancy and encoded as shown schematically in Figure 4-23 with $m = 200$, $k = r = 8$. The 1600 data bits labeled $D_1, \ldots D_{1600}$ are assigned to the positions shown in the data portion of the block. In this machine, there are eight 200-bit columns of data. The redundancy bits $R_1, \ldots R_{1600}$ are arranged in eight columns of 200 bits as shown. The data and redundancy bits in each row form a 16-bit subblock which is encoded using an eighth-degree polynomial capable of correcting any two errors within the subblock.

The system detects whether errors have occurred at random or in correctable bursts and corrects them accordingly. Random errors are corrected on a 16-bit subblock basis. When a correctable burst error has been detected, the decoder considers the 3200 bit block as one word that was encoded using a burst code polynominal of degree 1600. Thus, the data is always encoded in rows as shown in Figure 4-23 but is decoded in either of two ways depending upon the nature of the errors. In this system, the data and redundancy bits are interleaved so that the data bits of block i and the redundancy bits of block i-1 are alternated on the channel. This staggered interleaving eliminates the need for storing data in the encoder while the redundancy is being calculated, and doubles the burst length that can be corrected. Thus, the encoder must provide storage only for the redundancy associated with blocks i and i - 1. In operation, one storage area is used to interleave redundancy from block

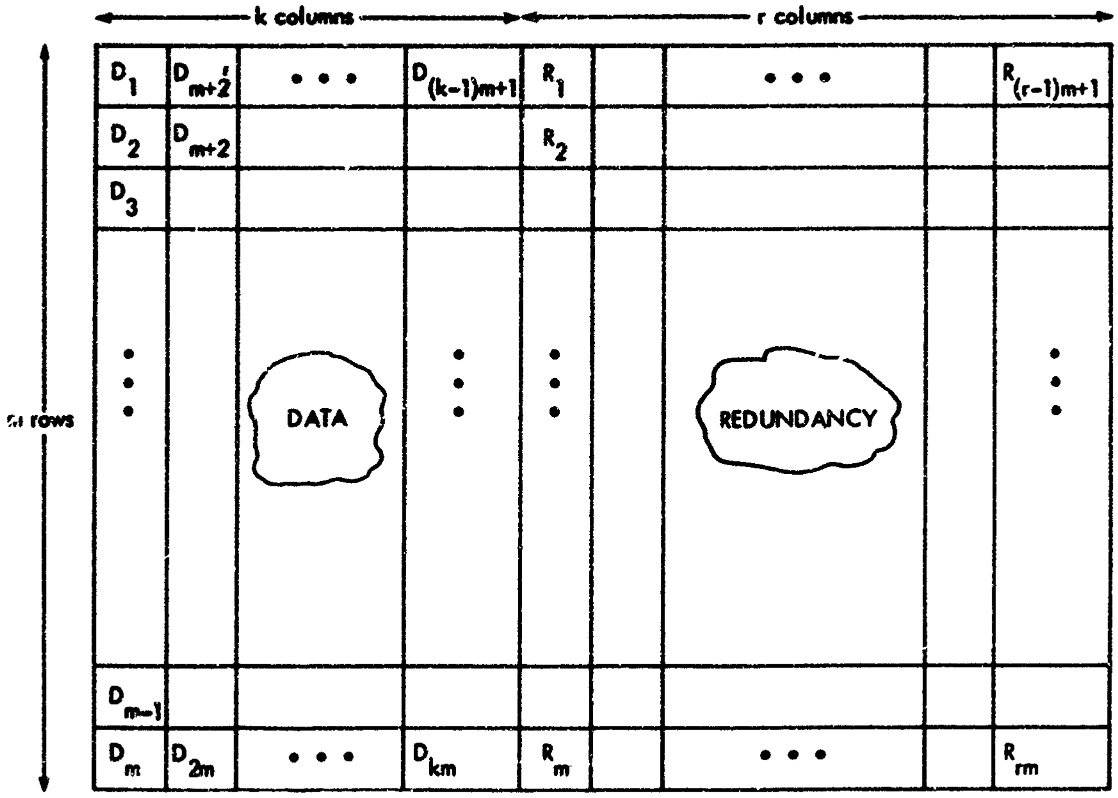


Figure 4-23. Block Format for Interleaved Variable Error Control Techniques; Block Length $N = m\,(k+r)$

i - 1 with data from block i while the other storage area is used to store the redundancy calculated from the data in block i.

This equipment has the capability of correcting bursts of 2800 bits. In the random correction mode, the system has the capability of correcting two errors in each interleaved 16-bit subword.

Initial block synchronization is established by searching for 200 consecutive error free subblocks and framing on them. Further block synchronization is then maintained by bit sync and counters. Since this equipment was built, an improved framing technique has been developed whereby only four consecutive interleaved subwords need be correctly received in order to obtain initial block synchronization.

The parameters used in this example are merely illustrative. The number of data bits or redundancy bits in each row as well as the number of rows are all arbitrary parameters for the system designer. The length, spacing, and density of expected bursts determine these parameters.

Section 5

EVALUATION OF ADAPTIVE MODULATION TECHNIQUES

The two techniques for adapting modulation parameters that appeared to have the most promise—variable bit duration and variable alphabet size—were selected for a detailed performance analysis. A considerable number of these systems, employing up to five different operating modes, were evaluated. To establish the extent to which they can improve the throughput and reliability provided by fixed techniques, they were analyzed over a wide range of channels exhibiting both slow and rapid fading.

5.1 CHANNEL MODEL

The model of the fading channel used in evaluating the adaptive modulation parameter communication systems is characterized by two statistical distributions, one describing the rapid fading in terms of a mean-value parameter and the other describing the distribution of that mean value, which is assumed to be slowly varying. Two extreme cases will be considered for the rapid fading, namely, a Rayleigh-distributed envelope (subsection 3.1.1), and a gaussian-distributed envelope (which is used here to approximate a Rician-distributed envelope—see subsection 3.1.2). In both cases, the mean value of the rapid fading distribution will be assumed to be a slowly-varying function of time with a log-normal distribution (subsection 3.1.4).

The threshold variable most readily available at a receiver for adapting to changes in the channel is the received signal power level, $P = S + N$, where S and N are the signal and noise powers respectively. Although the signal-to-noise ratio can be derived from the received signal power (see subsection 4.2), it is most convenient to work in terms of the direct measurement of received signal power.

For Rayleigh fading we have already seen (Equation 3-3) that the density function for the received signal power is given by

$$f(P) = (1/\overline{P}) \exp(-P/\overline{P}) \tag{5-1}$$

Since we will also be working with a log-normal distribution for $\overline{P}$, the analysis will be simplified by introducing the logarithmic variable $z = 10 \log_{10} P$, which is the decibel rating of the received signal power. Transforming Equation 5-1, we obtain the density function for z:

$$\phi(z) = \frac{\ln 10}{10\,\overline{P}} \exp\left(\frac{\ln 10}{10} z - \frac{10^{z/10}}{\overline{P}}\right) \tag{5-2}$$

where z ranges from $-\infty$ to $+\infty$. Further substitution of $Z = 10 \log_{10} \overline{P}$ gives

$$\phi(z) = \frac{\ln 10}{10 \cdot 10^{Z/10}} \exp\left(\frac{\ln 10}{10} z - \frac{10^{z/10}}{10^{Z/10}}\right) \tag{5-3}$$

Using the gaussian approximation to Rician fading (subsection 3.1.2) and converting Equation 3-13 to the logarithmic variable z, we obtain

$$\phi(z) = \frac{\ln 10 \exp(z \ln 10/10)}{20\sqrt{\pi \psi_0}} \left\{ \exp\left[-\frac{1}{2\psi_0} \left(\sqrt{2 \cdot 10^{z/10}} + \sqrt{2 \cdot 10^{Z/10}} \right)^2 \right] + \exp\left[-\frac{1}{2\psi_0} \left(\sqrt{2 \cdot 10^{z/10}} - \sqrt{2 \cdot 10^{Z/10}} \right)^2 \right] \right\} \tag{5-4}$$

for the density function.

It has already been shown (subsection 3.2.4) that the log of the average received power, $Z = 10 \log_{10} \bar{P}$ is normally distributed in the slow fading phenomenon. Thus we have the density function for Z given by

$$\gamma(Z) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2} (Z - \bar{Z})^2 \right] \tag{5-5}$$

where $\bar{Z}$ and σ are the mean and standard deviation of Z, respectively.

5.2 SELECTION OF MODE THRESHOLDS

The adaptive systems were designed to change modes on the basis of the expected average error probability, taking into account the rapid fading statistics, for particular values of the mean Z (in db) of the rapid fading. For a system with N modes, let the modes be so ordered that Mode 1 has the slowest rate (i.e., smallest probability of error for a given signal-to-noise ratio), Mode 2 is next slowest, etc. For each Mode i, $i = 2, \ldots, N$, let the maximum average bit error probability to be tolerated in that i^{th} mode be $P_{\theta_{i-1,\,i}}$. Then the threshold between Mode i - 1 and Mode i will be that Z, such that the error probability averaged over the rapid fading with mean Z equals $P_{\theta_{i-1,\,i}}$.

Thus, we find the value of Z such that

$$P_{\theta_{i-1,i}} = \frac{1}{2}\left(\frac{M_i}{M_i - 1}\right)\int_{-\infty}^{+\infty} P_{E_i}(z)\,\phi\,(z, Z)dz \qquad (5\text{-}6)$$

where M_i is the order of the alphabet in Mode i, $P_{E_i}(z)$ is the formula for symbol error probability in Mode i, which is related via (4-8) to bit error probability, and $\phi\,(z, Z)$ corresponds to $\phi\,(z)$ of Equations (5-3) or (5-4) with average signal power Z db. The value Z thus found is taken to be the threshold $\theta_{i-1,\,i}$ between Mode i - 1 and Mode i.

Let $P_i(Z)$ denote the right side of Equation (5-6). Then the desired threshold is that value of Z for which

$$P_i(Z) = P_{\theta_{i-1,\,i}} \,. \qquad (5\text{-}7)$$

Since $P_i(Z)$ is a monotonically decreasing function of Z, a binary search was used to find Z. For any value of P_θ of interest, it is reasonable to assume that the value of Z which satisfies Equation (5-7) lies between -5σ and $+5\sigma$. To begin the search, $P_i(Z)$ could be evaluated at an arbitrary number in the interval $(-5\sigma, +5\sigma)$, but for efficiency $Z = 0$ should be chosen since it is in the middle of the interval. $P_i(0)$ is compared to $P_{\theta_{i-1,i}}$. If $P_i(0) > P_{\theta_{i-1,i}}$, the value of Z which satisfies Equation (5-7) must lie in the interval $(0, 5\sigma)$ since $P_i(Z)$ is a monotonically decreasing function. The next value assigned to Z is $(1/2)(0 + 5\sigma)$ since 5σ is an upper bound and 0 is now known to be a lower bound for the Z satisfying Equation (5-7). If $P_i(0) < P_{\theta_{i-1,i}}$, the next value tried for Z is $(1/2)(0 - 5\sigma)$ since -5σ is a lower bound, and now 0 is known to be an upper bound for the Z which satisfies Equation (5-7). $P_i(Z)$ is evaluated at the Z thus obtained. Each subsequent value of Z is chosen according to the following rule:

$$Z = \frac{1}{2}(Z_p + U), \text{ if } P_i(Z_p) > P_{\theta_{i-1,i}} \tag{5-8a}$$

$$Z = \frac{1}{2}(Z_p + L), \text{ if } P_i(Z_p) < P_{\theta_{i-1,i}} \tag{5-8b}$$

where Z_p is the preceding value tried for Z, L is the lower bound found from previous trials, and U is the upper bound found previously. This process is continued until a value of Z is found such that

$$\left| P_i(Z) - P_{\theta_{i-1,i}} \right| \leq \epsilon . \tag{5-9}$$

A program was written to determine Z. For this program ϵ was chosen to be 10^{-6}. The value of Z thus obtained becomes $\theta_{i-1,i}$. This process is repeated for each pair of adjacent modes.

The function $P_i(Z)$ was evaluated by use of the gaussian quadrature with 50 points of evaluation.

5.3 AVERAGE THROUGHPUT

The fraction of time spent in any Mode i is given by

$$\int_{\theta_{i-1,\, i}}^{\theta_{i,\, i+1}} \gamma\,(Z)\; dZ \tag{5-10}$$

Hence, for an N-mode system, the average equivalent number of binary digits per second is given by

$$\begin{aligned} \overline{R} = & \int_{-\infty}^{\theta_{1,\,2}} R_1 \log_2 M_1 \;\gamma\,(Z)\; dZ \\ & + \int_{\theta_{1,\,2}}^{\theta_{2,\,3}} R_2 \log_2 M_2 \;\gamma\,(Z)\; dZ \\ & + \;.\;.\;.\; + \int_{\theta_{N-1,\,N}}^{\infty} R_N \log_2 M_2 \;\gamma\,(Z)\; dZ \end{aligned} \tag{5-11}$$

where R_i is the throughput in Mode i (Equation 4-6).

In programs for numerical computation, different methods of calculation were used depending on the number of modes. For one-mode systems, the average throughput was calculated by evaluating

$$\int_{-\infty}^{\infty} R \log_2 M \;\gamma\,(Z)\; dZ \tag{5-12}$$

with the lower and upper limits approximated by $\bar{Z} - 5\sigma$ and $\bar{Z} + 5\sigma$, respectively. If two modes were being considered, the average throughput was calculated by evaluating

$$\int_{-\infty}^{\theta_{1,\,2}} R_1 \log_2 M_1 \;\gamma(Z)\, dZ + \int_{\theta_{1,\,2}}^{\infty} R_2 \log_2 M_2 \;\gamma(Z)\; dZ \qquad (5\text{-}13)$$

by using the Laguerre quadrature after applying the following transformations:

$$\int_{-\infty}^{\theta_{1,2}} f(x)\, dx = \int_0^{\infty} f(-x + \theta_{1,\,2})\, dx \qquad (5\text{-}14a)$$

and

$$\int_{\theta_{1,\,2}}^{\infty} f\,(x)\; dx = \int_0^{\infty} f\,(x + \theta_{1,\,2})\; dx \qquad (5\text{-}14b)$$

For three or more modes the average throughput was calculated by

$$\int_{-\infty}^{\theta_{1,\,2}} R_1 \log_2 M_1 \;\gamma(Z)\, dZ + \int_{\theta_{1,\,2}}^{\theta_{2,\,3}} R_2 \log_2 M_2 \;\gamma(Z)\; dZ$$

$$+ \int_{\theta_{i-1,\,i}}^{\theta_{i,\,i+1}} R_i \log_2 M_i \;\gamma(Z)\, dZ + \ldots + \int_{\theta_{N-1,\,N}}^{\infty} R_n \log_2 M_n \;\gamma(Z)\, dZ$$

(5-15)

where N is the number of modes. Again, the transformations were made on the first and last integrals, which were integrated by the Laguerre quadrature. The other integrals were integrated by the gaussian quadrature method using fifty points.

5.4 AVERAGE ERROR PROBABILITY

Because a two-stage model of the fading is used, the error probability must be averaged first with respect to the rapid fading, and then with respect to the slowly varying mean of the rapid fading.

First the average number ϵ_i of equivalent binary errors per second in each Mode i is found:

$$\epsilon_1 = \frac{R_1 M_1}{2(M_1-1)} \int_{-\infty}^{\theta_{1,\,2}} \left[\int_{-\infty}^{\infty} P_{E_1}(z)\, \phi(z,\, Z)\, dz \right] \gamma(Z)\, dZ \quad (5\text{-}16a)$$

$$\epsilon_i = \frac{R_i M_i}{2(M_i-1)} \int_{\theta_{i-1,\,i}}^{\theta_{i,\,i+1}} \left[\int_{-\infty}^{\infty} P_{E_i}(z)\, \phi(z,\, Z)\, dz \right] \gamma(Z)\, dZ \quad (5\text{-}16b)$$

for $i \neq 1$, N.

$$\epsilon_N = \frac{R_N M_N}{2(M_N-1)} \int_{\theta_{N-1,\,N}}^{\infty} \left[\int_{-\infty}^{\infty} P_{E_N}(z)\, \phi(z,\, Z)\, dz \right] \gamma(Z)\, dZ \quad (5\text{-}16c)$$

The average equivalent binary error probability is then

$$\bar{p} = \frac{\epsilon_1 + \epsilon_2 + \ldots + \epsilon_N}{\bar{R}} \tag{5-17}$$

where $\bar{R}$ is the average throughput described in subsection 5.3.

For numerical computation the limits of $-\infty$ and $+\infty$ on the outer integral were approximated by $Z - 5\sigma$ and $Z + 5\sigma$ whenever they appeared, and the gaussian quadrature with fifty points of evaluation was used. For each of these fifty points, the inside integral with the limits approximated by -40 and $+40$ was evaluated at fifty points using the gaussian quadrature.

5.5 TRANSITION COSTS

Each transition from Mode i to Mode j is accompanied by a period of average duration t_a^{ij} during which the channel is in a new state (i.e., the threshold variable has crossed a threshold) but the system continues to operate in the old mode. This period results from inevitable delays in recognizing the occurrence of threshold crossings, conveying this information to the other terminal, and changing to the new operating mode. In addition, in some systems there will be an idling time of duration t_d^{ij} associated with each change of mode.

These transitional delays will alter the throughput and error rate calculations shown in subsections 5.3 and 5.4. The appropriate correction terms are derived below.

5.5.1 Throughput with Transitional Delays

For each transition from Mode i to Mode j, there is a loss in throughput equal to $t_d^{ij} R_j \log_2 M_j$ bits due to dead, or idle, time and a "loss" of $t_a^{ij} (R_j \log_2 M_j - R_i \log_2 M_i)$ bits due to operating in the old modes after the channel has changed. It should be noted that the second loss will actually be a gain if the new mode is slower. This gain is made, of course at the expense of additional errors.

If we let $N_{i,\,i+1}$ denote the number of times per second that the threshold $\theta_{i,\,i+1}$ between Modes i and Mode i + 1 is crossed in each direction (see Equations 3-34 and 3-35), then the average throughput (Equation 5-11) must be adjusted by the subtraction of the expression:

$$L = \sum_{i=1}^{N-1} N_{i,\,i+1} \Big[t_d^{i,\,i+1} R_{i+1} \log_2 M_{i+1}$$

$$+ t_a^{i,\,i+1} (R_{i+1} \log_2 M_{i+1} - R_i \log_2 M_i) \Big]$$

$$+ \sum_{i=2}^{N} N_{i,\,i-1} \Big[t_d^{i,\,i-1} R_{i-1} \log_2 M_{i-1}$$

$$+ t_a^{i,\,i-1} (R_{i-1} \log_2 M_{i-1} - R_i \log_2 M_i) \Big] \qquad (5\text{-}18)$$

The adjusted throughput rate is thus given by

$$R = \overline{R} - L \qquad (5\text{-}19)$$

5.5.2 Error Rates with Transitional Delays

Three correction terms are required to adjust the error rate $\overline{\rho}$ calculated from Equation 5-17. We can approximate the adjustment in error rate due to remaining in the old mode after the channel has crossed the threshold into a new state by using the average error probabilities for the two modes at the threshold between them. This adjustment is given for the transition from Mode i to Mode j by

$$\Delta_a^{ij} = t_a^{ij} \left(R_i \, P_{E_i}(\theta_{ij}) - R_j P_{\theta_{ij}}\right) \tag{5-20}$$

erroneous symbols per transition. This will be positive if Mode i has a higher error rate and negative otherwise. From $\overline{\rho}$ we must subtract the effect of

$$\Delta_d^{ij} = t_d^{ij} \, R_j P_{\theta_{ij}} \tag{5-21}$$

erroneous symbols per transition to account for errors that did not occur during the dead time at the transition, but which were included in the original computation without transition costs. Finally, the number E_{ij} of additional symbol errors expected at each transition from Mode i to Mode j due to synchronization or control problems must be added in.

Thus the total number of additional binary errors per second is given by

$$C = \sum_{i=1}^{N-1} \frac{1}{2} \frac{M_i}{M_i - 1} N_{i,\,i+1} (E_{i,\,i+1} + \Delta_a^{i,\,i+1} - \Delta_d^{i,\,i+1})$$

$$+ \sum_{i=2}^{N} \frac{1}{2} \frac{M_i}{M_i - 1} N_{i,\,i-1} (E_{i,\,i-1} + \Delta_a^{i,\,i-1} - \Delta_d^{i,\,i-1}) \tag{5-22}$$

and the adjusted output error rate is given by

$$\rho = \frac{\bar{R}\,\bar{\rho} + C}{\bar{R}} \tag{5-23}$$

5.6 QUANTITATIVE RESULTS

The first question considered was whether, in an idealized situation without transition costs, adapting the bit rate would improve the throughput and reliability over a fixed bit rate system. In conjunction with this, the choice of threshold probabilities $P_{\theta_{i,\,i+1}}$ was also investigated.

A low quality FSK tropospheric scatter link with an overall error rate of 0.0126 at 1000 bps was examined. For this channel, a Rayleigh fast fading envelope with Z, the log of the mean received power level, normally distributed with mean $\bar{Z}$ = 25 db and standard deviation σ = 10 db was used as a model. Five different bit rates, 125, 250, 500, 1000, and 2000, and four different threshold probabilities, 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2}, were considered. The results are given in Table 5-1.

Table 5-1. Performance of Variable Bit Rate Systems on Tropospheric Scatter Link

System	$P_{\theta_{i,i+1}}$	R $\frac{\text{bits}}{\text{sec}}$	ρ $\frac{\text{bits}}{\text{bit}}$
125 bits/sec. fixed	----	125	0.0023
250 bits/sec. fixed	----	250	0.0041
500 bits/sec. fixed	----	500	0.0073
1000 bits/sec. fixed	----	1000	0.0126
2000 bits/sec. fixed	----	2000	0.0209
125/250/500/1000/2000 adaptive	10^{-5}	152	0.0019
125/250/500/1000/2000 adaptive	10^{-4}	330	0.00090
125/250/500/1000/2000 adaptive	10^{-3}	862	0.00075
125/250/500/1000/2000 adaptive	10^{-2}	1547	0.00290

Several conclusions can be drawn from Table 5-1. First, in this idealized situation, adapting the bit rate offers simultaneous improvement of throughput rate and error rate. Second, the threshold probability criteria should not be too ambitiously selected. Choices like 10^{-5} and 10^{-4} call for considerably fewer errors over the short term than the slowest mode alone can provide in the long term. Examination of the actual thresholds $\theta_{i,i+1}$ found by the program indicated that with overambitious threshold probabilities the system rarely adapted into the faster modes.

These results are typical of those obtained with other sets of parameters. Increasing or decreasing the noise power or mean signal power or changing modulation schemes merely changed the error rates for the various transmission rates but did not alter the relative performance of the various rates or rate combinations. Even changing the fading model did not alter the conclusions. Table 5-2 shows similar results for an HF ionospheric reflection channel subject to the same slow fading but having gaussian rapid fading with $a = E/\sqrt{\psi_0}$ set equal to 3 and 10 and with the corresponding error rates at 100 bits per second equal to 0.0050 and 0.0033, respectively.

Table 5-2. Performance of Variable Bit Rate Systems on General HF Ionospheric Reflection Links

System	a=3		a=10	
	R	ρ	R	ρ
25 bits/sec.	25	0.0010	25	0.00063
50 bits/sec.	50	0.0023	50	0.0014
100 bits/sec.	100	0.0050	100	0.0033
200 bits/sec.	200	0.011	200	0.0082
400 bits/sec.	400	0.022	400	0.020
adaptive 25/50/100/200/400, $P_\theta = 10^{-2}$	289	0.0033	283	0.0051
adaptive 25/50/100/200/400, $P_\theta = 10^{-3}$	142	0.00062	87	0.00062

The investigation also considered the effects of using different numbers and spacings of modes. It was found that, although each increase in the number of modes resulted in improved performance, the step from a fixed, one-mode system to an adaptive two-mode system was the most significant. Furthermore, the two modes should be widely spaced in capability. This is demonstrated in Table 5-3 for the tropospheric scatter link considered above.

Table 5-3. Effects of Varying the Numbers and Definitions of Modes

System	$P = 10^{-2}$		$P = 10^{-3}$	
	R	ρ	R	ρ
125/250/500/1000/2000	1547	0.0029	862	0.00075
250/2000	1470	0.0029	790	0.0016
125/2000	1430	0.0024	704	0.00072
125/1000	817	0.0022	494	0.00085
125/250	240	0.0020	207	0.0015

Again, the results were similar with different channel parameters or with the variation in transmission rate achieved by increasing the order of the alphabet.

The results given above do not include the various losses due to delays in making transitions. If we assume that the tropo link under investigation is a 200-mile path and that service information is sent every 100 bit times (of the mode being employed), then the following results hold.

For slow fading at a rate of 0.0003 fades/sec there was no change in the throughput rate (measured to the nearest bit per second) or in the bit error rate (to three significant digits).

For slow fading at the rate 0.03 fades/second, the results obtained for threshold criteria $P_{\theta_{i,i+1}} = 10^{-2}$ are shown in Table 5-4.

Table 5-4. Effects of Transition Delays with $P_{\theta} = 10^{-2}$

System	Without Transition Delays		With Transition Delays	
	R	ρ	R	ρ
125 bits/sec	125	0.00226	125	0.00226
250	250	0.00413	250	0.00413
500	500	0.00733	500	0.00733
1000	1000	0.0126	1000	0.0126
2000	2000	0.0209	2000	0.0209
125/250	240	0.00199	240	0.00197
125/500	450	0.00198	447	0.00192
125/1000	817	0.00216	807	0.00205
125/2000	1430	0.00245	1403	0.00229
125/250/500/1000/2000	1547	0.00291	1544	0.00289
250/500	467	0.00325	466	0.00324
250/1000	843	0.00290	840	0.00286
250/2000	1470	0.00287	1458	0.00280

From this table it can be seen that the decrease in throughput is slight, ranging from 0.2 percent for the five-mode system and the closely spaced two-mode system to 1.9 percent for the most widely spaced two-mode system. These slight throughput losses were accompanied by slight decreases in the error rate ranging from 0.3 percent for five modes to 6 percent for the 125/2000 system.

The corresponding results for $P_{\theta_{i,i+1}} = 10^{-3}$ are shown in Table 5-5.

Table 5-5. Effects of Transition Delays with $P_\theta = 10^{-3}$

System	Without Transition Delays		With Transition Delays	
	R	ρ	R	ρ
125/250	207	0.00151	206	0.00151
125/500	328	0.00109	323	0.00109
125/1000	494	0.000850	481	0.000843
125/2000	704	0.000724	677	0.000711
125/250/500/1000/2000	862	0.000750	858	0.000746
250/2000	790	0.00159	778	0.00160

The case where the slow fade rate was 0.3 fades per second was also examined. This is faster than the duration normally considered to be slow fading. Here the throughput loss was more noticeable, but still tolerable. Some examples are shown in Table 5-6 for $P_{\theta_{i,i+1}} = 10^{-2}$ along with corresponding figures from Table 5-5.

Table 5-6. System Performance with Several Slow Fade Rates

System	Performance with Transition Delays			
	$f_b = 0.03$		$f_b = 0.3$	
	R	ρ	R	ρ
125/250/500/1000/2000	1544	0.00289	1520	0.00269
125/250	240	0.00197	236	0.00177
125/500	447	0.00192	424	0.00133
125/1000	807	0.00205	722	0.000937
125/2000	1403	0.00229	1163	0.000570

Additional runs with different sets of parameters were quite similar and substantiated the following conclusions:

1. Varying the bit duration or the number of bits per symbol can provide significant improvement in both the throughput rate and error rate in real fading channels.

2. As the number of modes becomes larger the improvement becomes greater.

3. In view of the added hardware and control logic required for extra modes, and in view of the fact that the most significant improvement accompanies the addition of a second mode, two-mode systems appear to be the most practical.

4. The transmission rates in the two modes should be as widely separated as system constraints allow.

5. The thresholds between these modes should be set low enough so that the system can spend as large a fraction of the time in the high-speed mode as possible.

6. For slow fading within the range of interest, the loss in throughput due to propagation delays and delays in conveying service information has no significant relation to the improvement obtained over fixed systems.

7. The throughput losses are generally accompanied by insignificant decreases in the error rate.

8. The throughput losses are smaller for multimode systems than for two mode systems and increase as the separation of the modes increases.

Section 6

EVALUATION OF ADAPTIVE ERROR CONTROL TECHNIQUES

Adaptive error control techniques are sensitive to very fine details in the serial occurrence of errors. Average numbers of errors per block or distributions of burst lengths do not provide sufficient statistics for an analytical evaluation of adaptive error control techniques. The additional factors which must be considered, namely the reaction time of the system, the behavior of control information, the extent to which the capabilities of the code are exceeded, and the net effects of such occurrences do not lend themselves to an analytically tractable solution. Moreover, such a solution would give little insight into the problems of adaptive system design and performance. Accordingly, it was decided to evaluate the fixed block length variable redundancy technique and the variable block length fixed redundancy technique by computer simulation of both the channel and the adaptive technique. A meaningful analysis of this technique was thus obtained.

The fixed block, variable utilization of fixed redundancy technique presented a special situation. An operating prototype of such a system had been built and was available in the IBM Engineering Laboratory for testing. Since modifying the design was not feasible within the scope of this project, only the one set of adaptive parameters built into the device was used. The device was tested against serial error tapes.

6.1 SIMULATOR FOR ADAPTIVE CODING TECHNIQUES (SACT)

6.1.1 Introduction

SACT is an IBM 7090/94 FORTRAN II computer program designed to simulate and evaluate adaptive coding techniques. The program, depicted in Figure 6-1, consists of two major portions—the channel portion and the transmission/reception portion. Either portion may be modified at will without affecting the other portion. Figures 6-2a through 6-2e show the logic of a transmitter/receiver pair in sending, receiving, and analyzing blocks.

6.1.2 Channel Model

A wide variety of channel models can be used, including random errors, statistical fading models, or actual serial error bit streams obtained from experiments on transmission links. As indicated by Figure 6-1, the channel models for Channel A and Channel B can be different. Initially, however, the two channels will both use the same fading model, but with different random number sequences. The occurrences of errors then will be statistically independent.

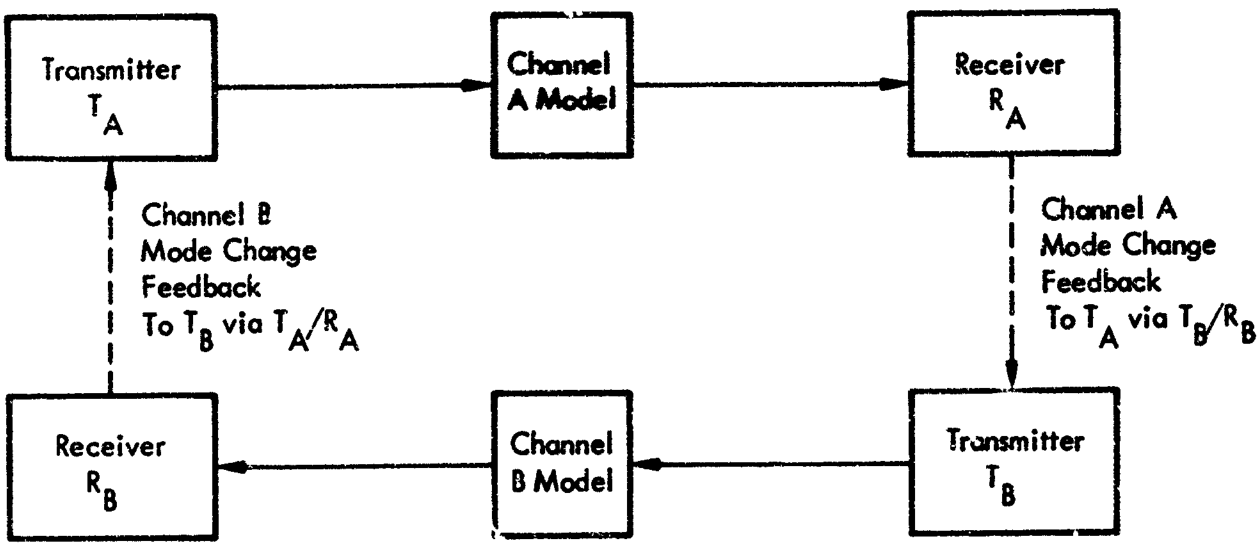


Figure 6-1. SACT Organization

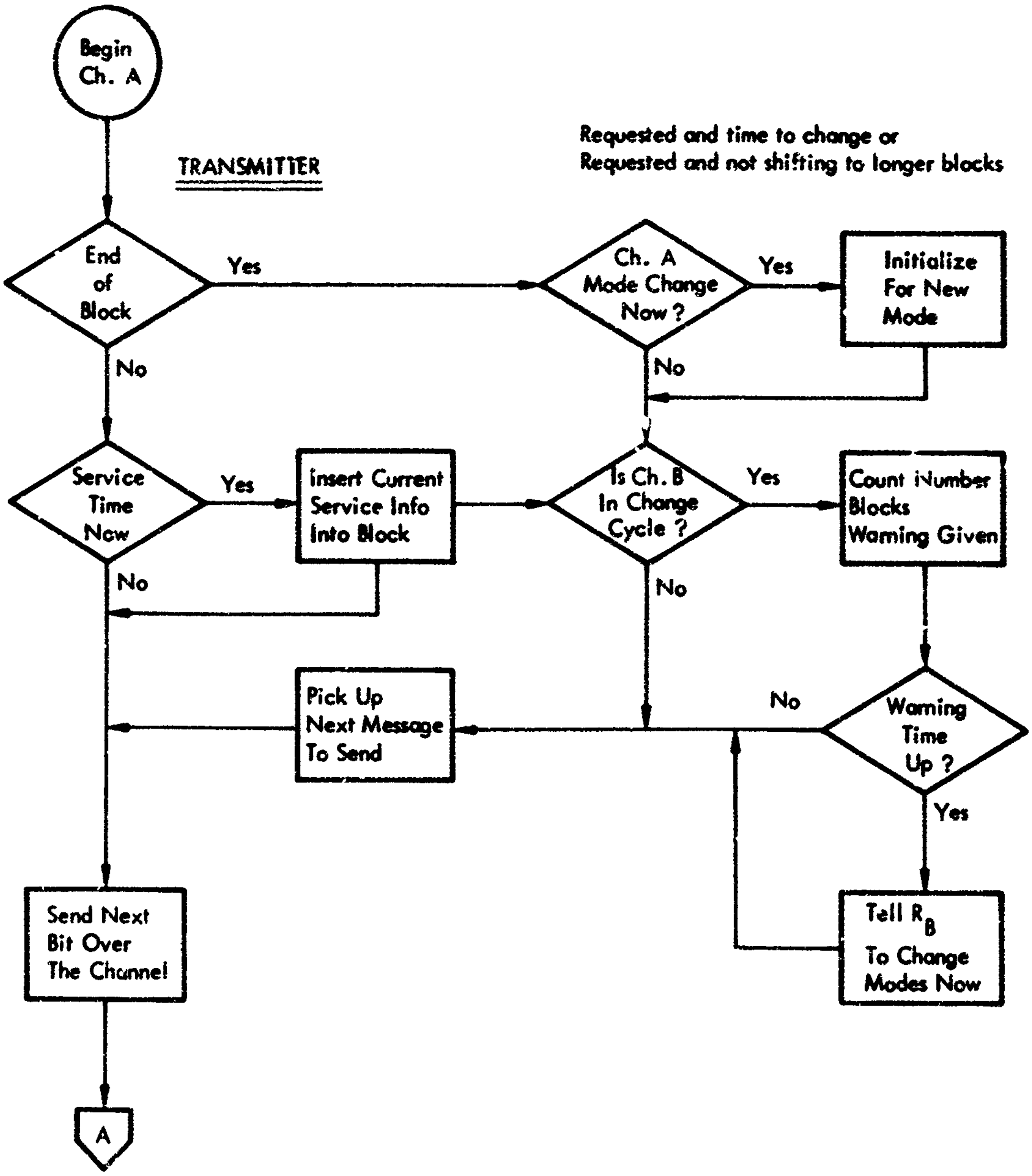


Figure 6-2a. SACT Flow Diagram

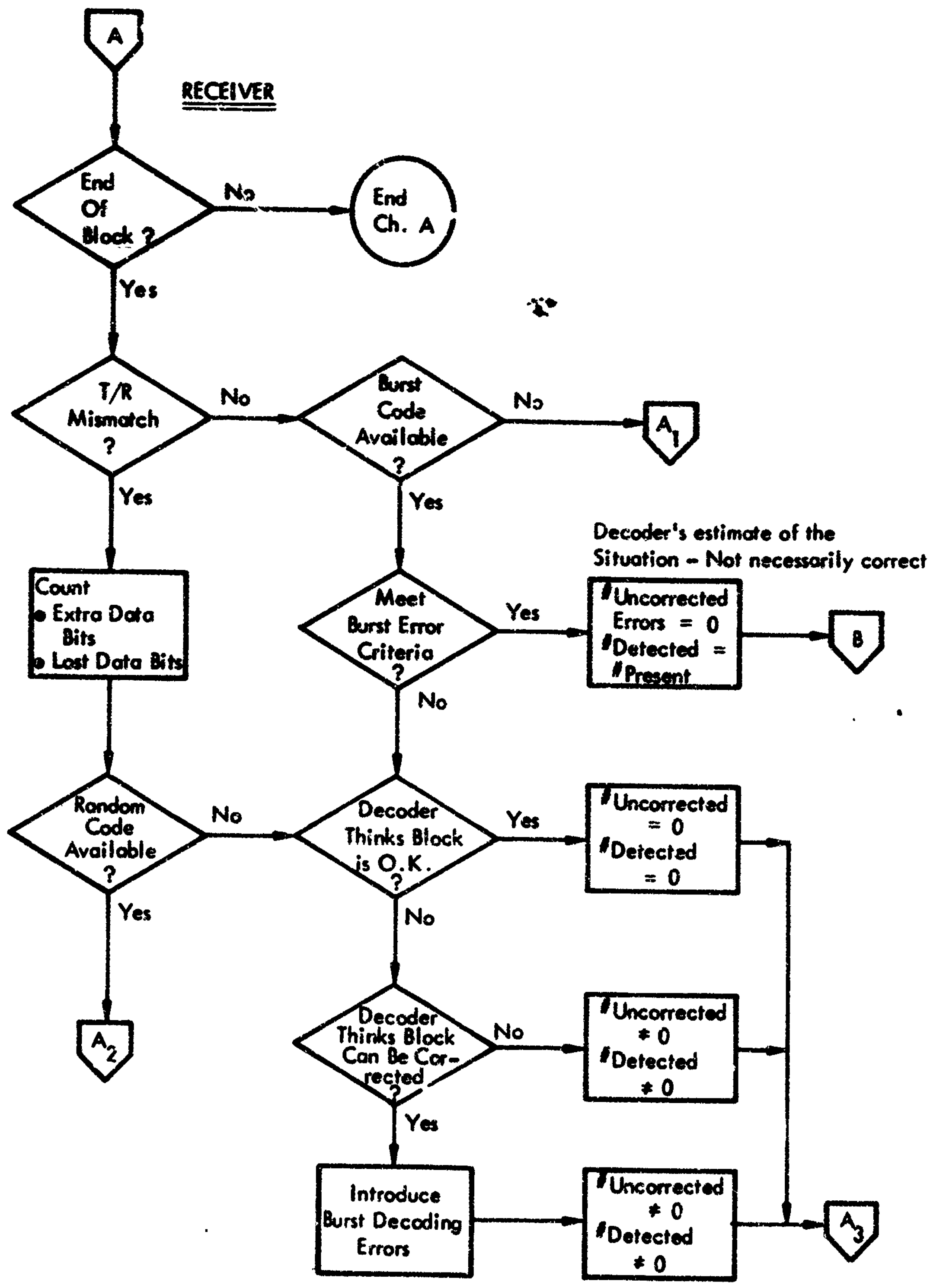


Figure 6-2b. SACT Flow Diagram

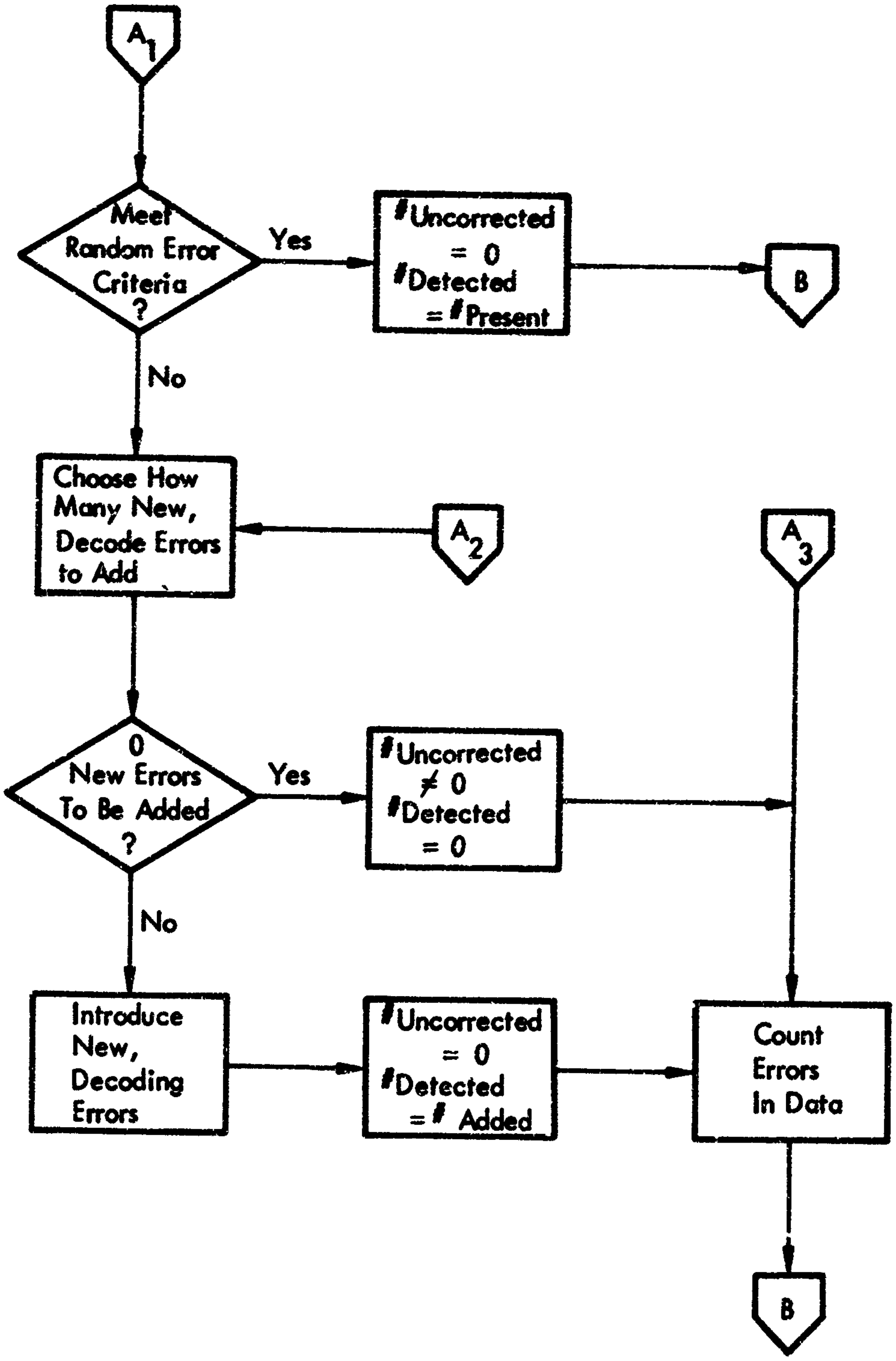


Figure 6-2c. SACT Flow Diagram

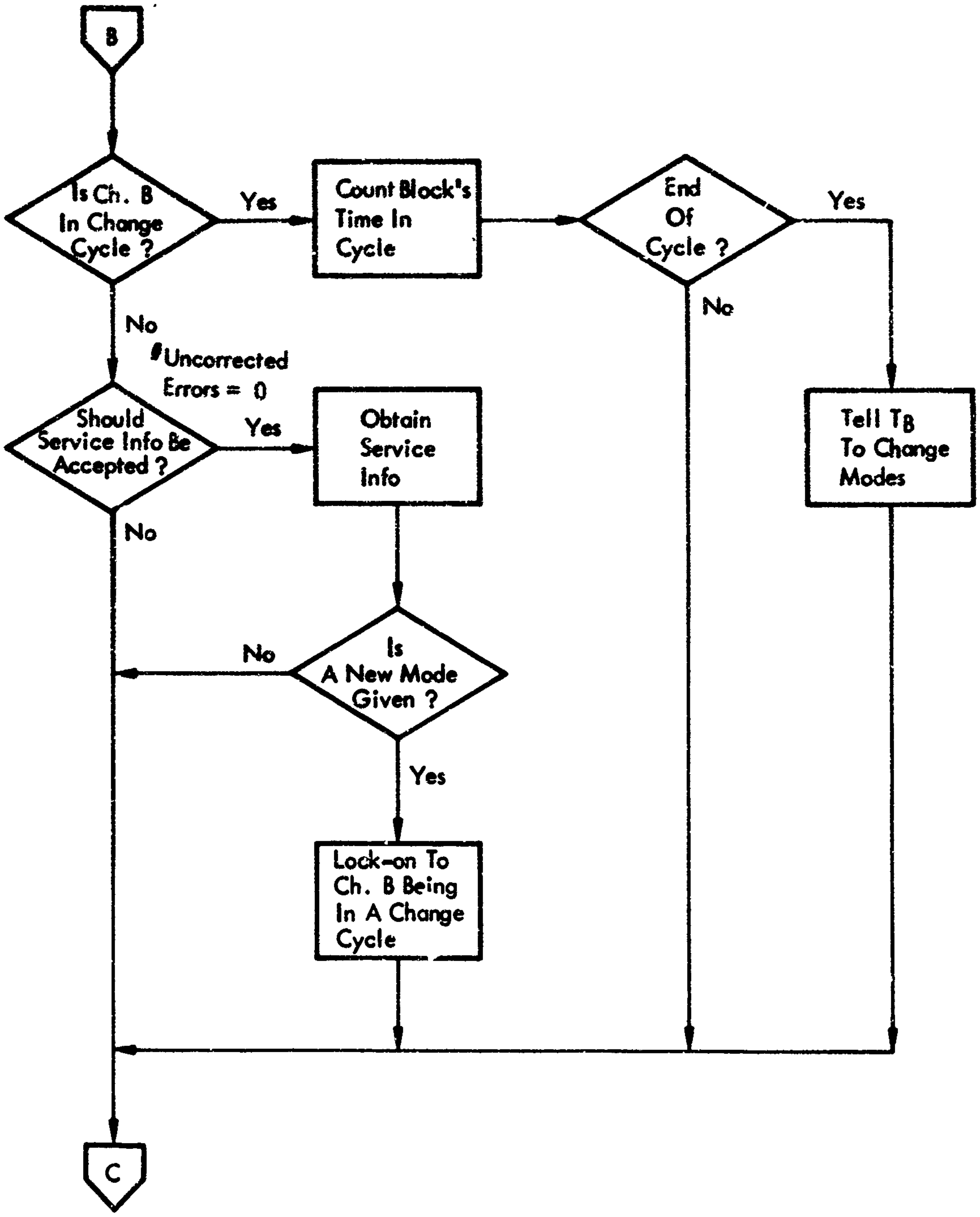


Figure 6-2d. SACT Flow Diagram

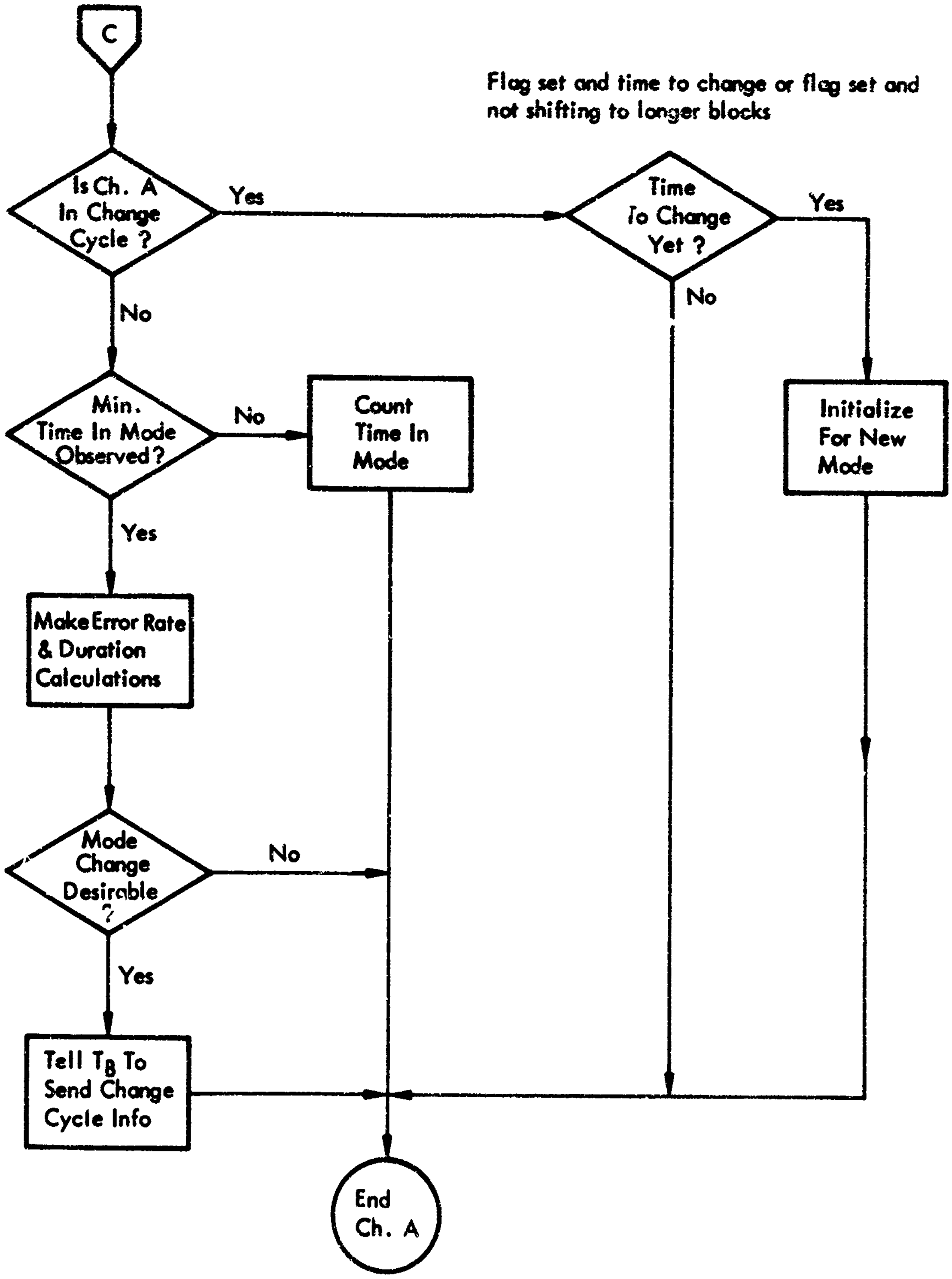


Figure 6-2e. SACT Flow Diagram

6.1.3 Transmission/Reception Model

The transmission/reception portion of SACT incorporates several genera concepts in the organization of the adaptive coding logic. Such concepts, ther fore, become limitations to the modeling of techniques. Basically, the progr: assumes that data is transmitted in blocks whose format is determined by the mode of transmission. Each channel operates independently and attempts to maintain a high throughput rate at a permissible error rate. Redundancy bits are used in each block to detect and correct erroneous bits. If bits in error exceed the capability of the code, some erroneous data will be accepted by the receiving terminal. The simulator does not have retransmission capability.

Every block has three portions: data, redundancy, and service message. The service message has two sections, the "mode" and the "warning time." The mode indicates the current mode in which the other channel is supposed t be whenever the warning time is zero. If the warning time is a number t not equal to zero, the mode indicates the mode to which the other channel should change, t blocks from that time.

Alternate block formats may be implemented to simulate other schemes, e.g., a block where data bits and redundancy bits are interleaved such that the block can be thought of as several identically formatted subblocks.

For the purposes of SACT, codes have been classified by the number of random errors or the maximum burst span that can be corrected. Also described is the statistical effect of the decoder when it receives a block that exceeds the correction capabilities of the code. By completely describing the codes in this fashion, the simulator does not waste time performing the polynomial coding/decoding operations.

SACT assumes that both channels operate at the same bit rate and that bit synchronism exists. However, the two channels act independently in adapting to error situations by changing modes.

The simulator does not assume that the transmitter and receiver on a channel are using the same mode. The mismatched mode situation can arise when a receiver is unable to properly communicate a desired change to the other transmitter due to a high error rate on the other channel. During the mismatched state of the system (if it occurs), the simulator will maintain statistics peculiar to the state. Included in these statistics are the number of data bits lost and the number of non-data bits thought to be data bits.

Restoration of matched transmitter/receiver modes depends upon the transmitter learning that it should be in the new mode and the receiver synchronizing to the new mode of transmission.

6.1.4 Input

Data provided to the simulator consists of several system parameters and a description of the modes to be used in the simulation. System parameters include

Length of time to simulate, in seconds

Channel bit rate, in bits per second

Initial offset in time of the beginnings of the first blocks in Channel A and Channel B.

Criteria as to how many errors a block can have and still use the service information

Parameters describing the channels.

The parameters that uniquely describe a mode are

Mode number, 1 to n for an n-mode system

Block length

Redundancy length

Minimum time one must remain in that mode

Warning time required for transitions from that mode

Span of burst errors that can be corrected

Number of random errors that can be corrected

Lowest error count desired without shifting to a higher mode

Number of blocks (x out of y) that must have an error count lower than the above to actually shift to the next higher numbered mode.

Highest error count that will be tolerated without shifting to the next lower numbered mode

Number of blocks (a out of b) that must have an error count higher than the above to shift to the next lower numbered mode

6.1.5 Output

The following raw statistical measures are provided in the output at the end of a run, for the system as well as for each channel separately:

Length of time simulated

Number of data bits transmitted

Number of data bits received

Number of erroneous data bits accepted

Number of data bits lost due to mode mismatch

Number of non-data bits accepted as data during mismatch conditions

Number of blocks sent in each mode

Number of transitions between modes; $M_i \rightarrow M_{i+1}$ $(1 \leq i < N)$, and $M_i \rightarrow M_{i-1}$ $(1 < i \leq N)$.

6.2 RESULTS FOR VARIABLE BLOCK LENGTH, FIXED REDUNDANCY

A considerable number of runs of variable block length, fixed redundancy systems were made against simulated tropo and HF links and serial error tapes for a tropo link obtained from the National Bureau of Standards. The results were discouraging for proponents of this technique, although they were readily explainable.

In summary, adapting the block length while keeping the redundancy fixed does not perform as well as a well-chosen fixed coding technique. The reason is easily seen. These channels are subject to long bursts. The way that long bursts are corrected with standard coding techniques requires long blocks with at least twice as many redundant bits as the length of burst to be corrected. This can be done with a long block code or its equivalent in interleaved short codes or recurrent codes. Thus the slow, burst-correcting mode, even if it has only 50 percent data, will have a block length exceeding four times the maximum correctable burst and a redundancy length twice the maximum correctable burst. This situation imposes severe restrictions on the fast, good channel mode. Since the redundancy is fixed from mode to mode, the good mode has at least twice the maximum correctable burst length in redundancy. Since it is to be more efficient, its block length must exceed the slow mode block length and hence exceed four times the maximum correctable burst.

The consequences render the system ineffective. The duration of the error-causing portions of fades is now less than one fourth of the block lengths. The mean time before the receiver can recognize that a burst has occurred is at least twice the duration of the burst. To this already hopeless delay must be added the time required for the transmitter to receive the request and complete the block then in progress before the first block in the new mode can be sent. The result is that the duration of the new channel state is often shorter than the reaction time of the system.

The newly developed statistical burst correcting codes require less redundancy than the standard techniques (see subsection 4.6.3). Even if these techniques are used, the conclusion is unchanged. The mean reaction time will still exceed the burst duration.

If the block length is not so long that multiple bursts occur in a single block, the fast mode block code corrects the burst. But then the system switches to a slow mode and the throughput is reduced. In other words, the fast mode alone outperforms the adaptive system. As an example from the simulation, consider a tropospheric scatter link whose slow fade rate has mean 0.016 cps, and fades 12 db below the median (with σ = 3db) have a mean duration of one second. The overall bit error rate is 5.5×10^{-4}.

Transmission is at 750 bits per second. The following results illustrate the points made above.

Table 6-1. Performance of Variable Block Length, Fixed Redundancy Systems

Block Length	Redundancy Length	Burst Capability	Throughput	Uncorrected Error Rate
3800	1900	950	0.500	0
7600	1900	950	0.750	0
3800/7600 adaptive	1900 (each mode)	950 (each mode)	0.703	0
3800/15200 adaptive	1900 (each mode)	950 (each mode)	0.809	0
15200	1900	950	0.875	0

In the 3800/7600 system, 88 percent of the data was sent in the faster mode, and in the 3800/15200 system 90 percent was sent in the faster mode.

A similar channel with deeper and more persistent fades such that the error rate was 10^{-3} gave the same fixed-mode system results. In the 3800/7600 system, 86 percent of the data was sent in the faster mode and the throughput was reduced to 0.690. In the 3800/15200 system, 87 percent of the data was sent in the faster mode and the throughput was 0.790. In one run of the 15200 fixed system a double burst caused an uncorrected error rate of 7×10^{-6}, and in one run of the 3800/15200 system a double burst caused an uncorrected error rate of 9×10^{-6}. Longer blocks were not investigated.

Against severe HF models with higher error rates resulting from shorter but more frequent fades, fixed error control did not fare as well as with the tropo models, but the uselessness of adapting the block length was still evident.

6.3 RESULTS FOR FIXED BLOCK LENGTH, VARIABLE REDUNDANCY

Fixed block length, variable redundancy systems were simulated in tropo and HF channels. The results indicated that this adaptive technique was of little value. The reasons were similar to those for variable block length, fixed redundancy.

The slow mode must have a large block length in order to contain the large number of redundant bits required to correct the long bursts. This means that the low redundancy, high throughput mode must also have a long block length (since fixed block length systems are being considered here). Then the time required to recognize the presence of a burst exceeds the length of the burst and it is too late to adapt.

Using the tropo link introduced in subsection 3.2 with transmission rate 750 bits per second and overall error rate 5.5×10^{-4}, the results of Table 6-2 were obtained. These results show that bursts almost always occur during the low redundancy mode and cease before the effective mode can begin.

Table 6-2. Performance of Fixed Block Length, Variable Redundancy Systems

Block Length, Redundancy	R	ρ
(3800, 1900) fixed	.500	0
(3800, 1900)/(3800, 40) adaptive	.963	5.4×10^{-4}
(7600, 1900) fixed	.750	0
(7600, 1900)/(7600, 40) adaptive	.968	5.0×10^{-4}

Other runs with higher or lower quality links produced the same results. Only when the lower redundancy mode alone could handle the bursts did substantial improvement result. At that point, however, the lower redundancy mode had high redundancy and there was nothing to be gained by switching, after the burst had been corrected, to a less efficient mode. Using a moderate redundancy mode and a high redundancy mode resulted in an improved, but nevertheless unsatisfactory system. For example, Table 6-3 shows results for a tropo link with an error rate of 9.5×10^{-4} resulting from long deep fades occurring on the average of once every 60 seconds.

Table 6-3. Comparative Performance of a Fixed Block Length, Variable Redundancy Technique on a Typical Tropo Link

Block Length, Redundancy	R	ρ
(3600, 1800) fixed	0.500	0
(3600, 900) fixed	0.750	4.0×10^{-4}
(3600, 1800)/(3600, 900) adaptive	0.693	2.6×10^{-4}

6.4 LABORATORY TEST OF ADAPTIVE DECODER

Several reels of analog tape containing, among other measurements, errors recorded on a 380-mile tropospheric scatter link between San Juan, Puerto Rico, and Grand Turk Island, British West Indies, by the National Bureau of Standards were obtained. These were the only serial error patterns that could be obtained for fading channels. Unfortunately the link was of extremely low quality and was not typical of other tropo links. The average error rate ranged from 0.002 to 0.05 over five-minute intervals.

To test the adaptive decoder in this channel, it was necessary to convert the analog information to a digital error stream and perturb an artificial data stream accordingly. A detailed description of the procedure and equipment is given below.

6.4.1 Test Equipment and Procedures

The signals from the analog tape recorder were applied to pin 3 of a Fairchild μA 710 high-speed differential comparator as shown in Figure 6-3. Typically, this signal appears as shown in Figure 6-4. A variable reference level was applied to pin 2 of the differential comparator. The relation between the two voltages is shown in Figure 6-4. The differential comparator produces a positive level whenever the voltage on pin 3 is negative with respect to the voltage on pin 2 as shown in Figure 6-4. The analog pulses that appear at the output of the tape recorder are nominally 2 volts peak-to-peak and are approximately 10 μs in duration.

The impulses were recorded at a tape speed of 2.4 inches per second. The tape was played back at a tape speed of 7.5 inches per second, resulting in a speedup of 7.5/2.4. The error impulses were obtained from a 2400-bit-per-second data train. Thus the tape playback represents the same error sequence at 7500 bits per second.

Referring to Figure 6-3, the pulses appearing at the output of the differential comparator are fed into a level converter circuit. This circuit accepts a binary waveform that is either at zero volts or +5 volts and inverts and converts this waveform so that it is either at -6 volts or zero volts.

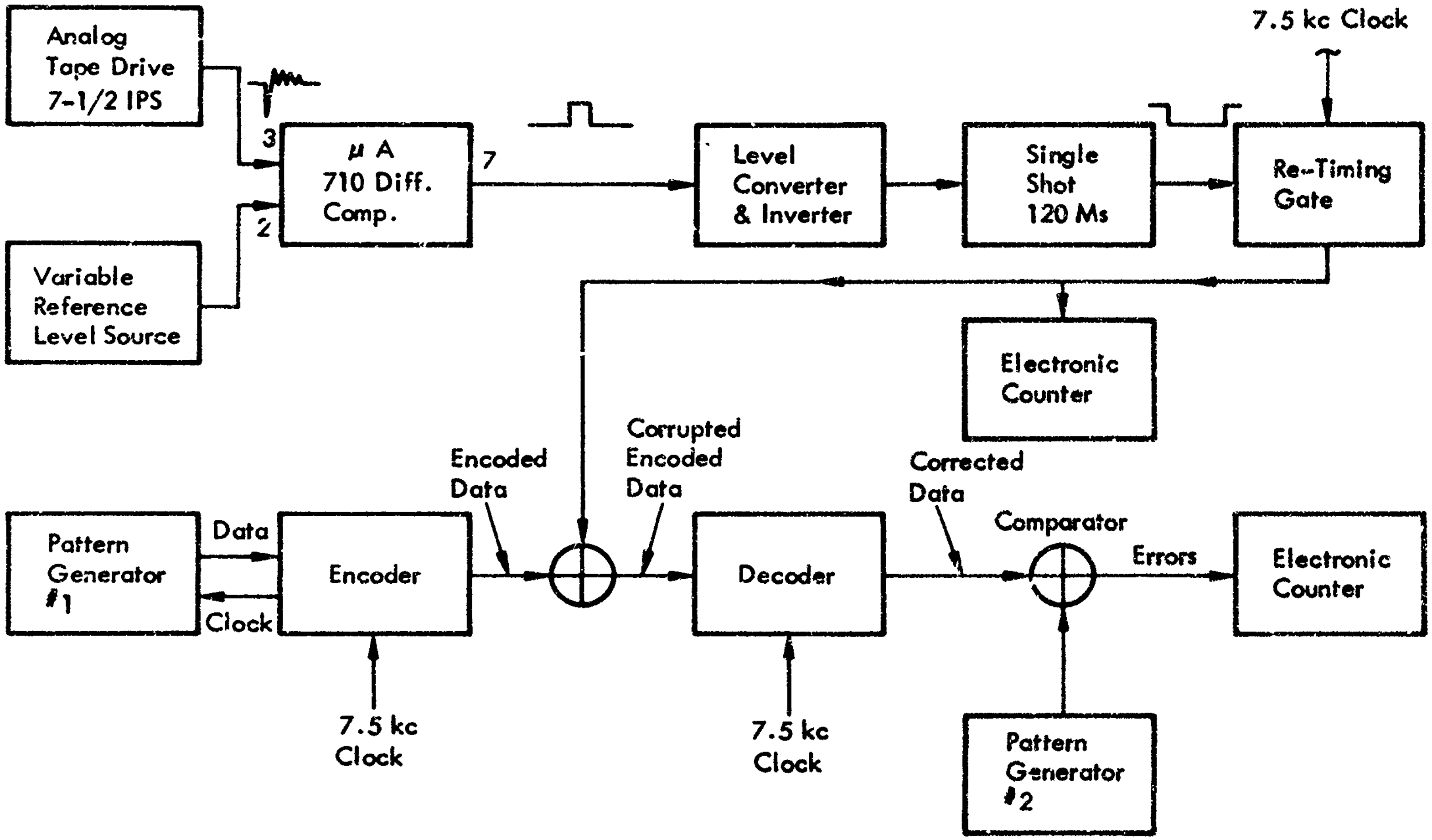
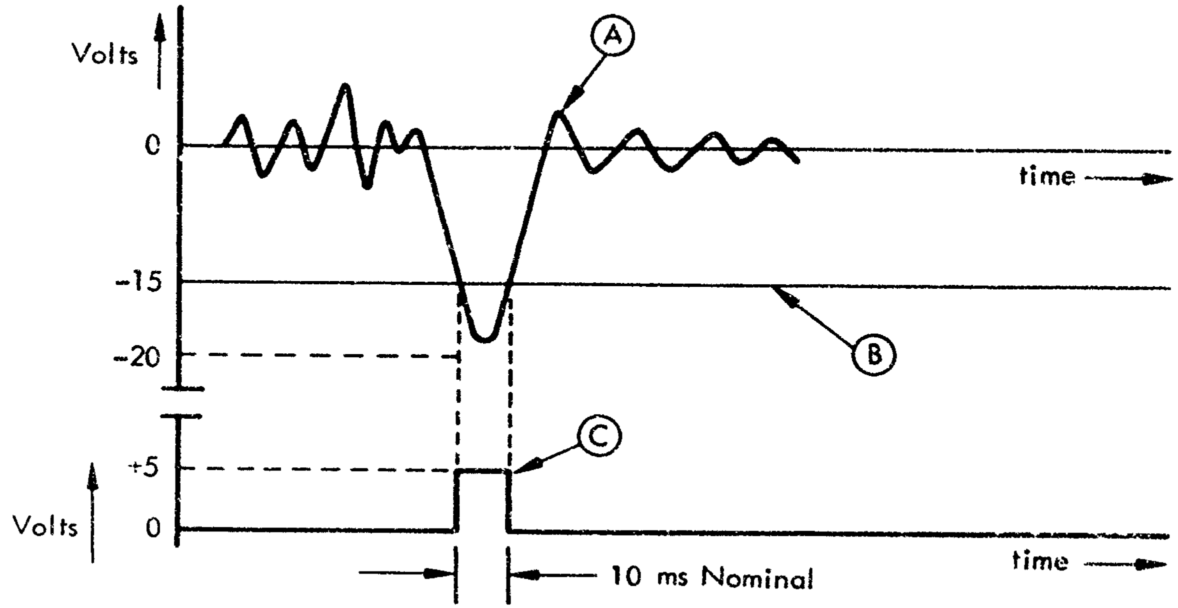


Figure 6-3. Error Pattern Recovery and Injection System



A. Output of analog tape recorder
B. Reference level
C. Output of high speed differential comparator

Figure 6-4. Differential Comparator

The output of the level converter and inverter drives a single shot circuit that produces a negative pulse 120 μs in duration. This circuit acts as a "pulse stretcher." The output of the single shot is used to condition a gate that is strobed by a clock waveform at a frequency of 7.5 kc. The purpose of this gate is to re-time the error pulses recorded on the tape. When a clock pulse and the output of the single shot coincide, an error pulse in phase with the system clock is generated. All error pulses are counted using an electronic counter. At this point the errors are ready to be injected into the channel.

A pattern generator serves as the data source. It generates a pseudo-random data stream that repeats every 511 bits. This data stream is fed into an encoder where it is encoded and transmitted to a decoder. The error pattern is added mod-2 to the encoded data by means of an exclusive-or logic circuit. Thus an error pulse at the output of the re-timing gate will cause the bit that is passing through the mod-2 adder to be inverted, thereby injecting an error into the encoded data stream.

The corrupted data is then passed on to the decoder where an error detection and correction process is performed. The output of the decoder is in the form of corrected data. This data is compared with the original data on a bit-by-bit basis. A second pattern generator, synchronized to the first, is used to generate a duplicate data stream for comparison purposes.

The output of the exclusive-or logic represents errors in the data stream generated by the decoder. These errors are counted using an electronic counter.

6.4.2 Test Results

The tropospheric scatter link error tapes obtained from the National Bureau of Standards were run through the operating model of the adaptive decoder using the test equipment and procedures described in subsection 6.4.1. The decoder used the theoretical principles described in subsection 4.6.3 with design parameters as indicated in subsection 4.7.2.3. It was capable of correcting 2800-bit bursts with a guard space of 4700 bits in the burst mode and 2 in 16 random errors in the random error correction mode, with switching between modes automatically controlled by the decoder.

The tapes had three usable 35-minute runs. Ambiguities in the locations of start and end of records limited experiments to 33 of the 35 recorded minutes in each run. Unfortunately, the quality of the link measured was extremely bad. For practical purposes it ranged from unusable to barely usable. The results for the entire reels are summarized below.

Run		Tape Data		Output Data	
NBS #	Bits	Errors	Rate	Errors	Rate
12-49	4,950,000	23,806	0.0048	4,140	0.00084
12-47	4,950,000	26,200	0.0053	1,554	0.00031
7-28	4,950,000	118,867	0.024	22,432	0.0045

The performance of the tropo link and the adaptive decoder for selected 450,000 bit segments is shown below.

Run	Errors In		Errors Out	
	Number	Rate	Number	Rate
12-49	881	0.0020	16	0.000036
12-49	427	0.0009	2	0.000004
12-49	6510	0.014	1558	0.0035
12-47	2399	0.00[illegible]	118	0.00026
12-47	4692	0.010	222	0.00049
12-47	1389	0.0031	28	0.000062
7-28	6808	0.015	486	0.0011
7-28	17596	0.039	5975	0.013

This indicates the variability of the link as well as the tremendous length of the bursts. One burst exceeding 12 seconds and containing more than 12,000 errors was noted. The improvement was several orders of magnitude in the link's better intervals, i.e., when the error rate was less than 0.005. There was improvement in all cases, including the three-fold improvement when the channel error rate was 0.039 for a three-minute period.

Section 7

CONCLUSIONS

A considerable number of adaptive techniques exist which are both feasible to implement and capable of providing significant improvements in the performance of digital data transmission systems. Techniques that vary either modulation parameters or error control parameters are included.

Among the variable modulation parameter systems, those that adapt either the bit rate or the alphabet size to changing channel conditions have the most potential. Typically, in a realistic fading channel using diversity reception, the error rate for a given average data throughput rate can be improved by one order of magnitude by adapting either of these parameters. In designing such a system, two different bit rates (or alphabet sizes) are sufficient. The gains from additional complexity are slight. The two operating modes should differ as much as possible, with the system constraints and adaptive thresholds chosen so that the system operates in the faster mode as much as possible. The control logic must be carefully designed to avoid loss of transmitter-receiver coordination.

Error control systems that vary the redundancy within a fixed block length or vary the block length while maintaining fixed redundancy cannot react quickly enough to degradations in channel performance to be worthwhile. Error detection with retransmission offers the highest throughput and reliability when feedback is used.

For applications without feedback there is a new error control technique in which the encoder operates in a fixed mode but the decoder monitors the channel and adapts its usage of the redundancy to correct either random or clustered errors. This adaptive feature gives this technique more capability to improve system performance than could be obtained from a random-correction-only code or from a burst-correction-only code. The magnitude of the improvement remains to be determined from on-line tests in real channels.

REFERENCES

1. F. P. Corr and A. H. Frey, Jr., "Error Control Through Coding," Volume VII, "Adaptive Error Control Systems," RADC-TDR-64-149, July 1964.

2. G. Lieberman, "Adaptive Digital Communication for a Slowly Varying Channel," Trans. AIEE, March 1963, pp 44-51.

3. S. Reiger, "Error Rates in Data Transmission," Proc. IRE, May 1959, pp 919-920.

4. S. M. Sussman, "Simplified Relations for Bit and Character Error Probabilities for M-ary Transmission over Rayleigh Fading Channels," IEEE Trans. on Communication Technology, Dec. 1964, pp 207-209.

5. G. Turin, "The Asymptotic Behavior of Ideal M-ary Systems," Proc. IRE, Jan. 1959, pp 93-94.

6. G. Axelby, E. Osborne, "Transmitter Power Control in Two-Way Communications System," IRE Transactions on Communication Systems, CS-8, March 1960, pp 48-56.

7. J. G. Lawton, "Investigation of Digital Data Communication Systems," Cornell Aeronautical Laboratory, RADC-TDR-61-58, Jan 3, 1961.

8. D. P. Harris, "Communication Through Random Multipath Media," Stanford Electronics Laboratory Technical Report No. 1002-2, April 1962.

9. B. Harris, A. Hauptschein, K. Morgan, L. Schwartz, "Binary Decision Feedback Systems for Maintaining Reliability Under Condition of Varying Strength," Proc. NEC, Oct. 1957.

10. R. J. Benice and A. H. Frey, Jr., "Comparisons of Error Control Techniques," IEEE Trans. on Communication Technology, Dec. 1964, pp 146-154.

11. R. Chien and D. Tang, "Error Control Through Coding," Volume III, "Variable Redundancy Codes," RADC-TDR-64-149, July 1964.

12. A. H. Frey, Jr.. "Message Framing and Error Control," IEEE Trans. on Military Electronics, April 1965, pp 143-147.

13. R. J. Benice and A. H. Frey, Jr., "An Analysis of Retransmission Systems," IEEE Transactions on Communication Technology, Dec. 1964, pp 135-145.

14. W. B. Davenport and W. L. Root, "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill, 1958.

15. S. O. Rice, "Mathematical Analysis of Random Noise," Selected Papers on Noise and Stochastic Processes, N. Wax Ed., Dover, 1954.

16. R. W. E. McNichol, "The Fading of Radio Waves of Medium and High Frequencies," Proc. IEE, Oct. 1949, p. 517.

17. K. Norton, L. E. Vogler, W. V. Mansfield, P. J. Short, "The Probability Distribution of the Amplitude of a Constant Vector Plus a Rayleigh Distributed Vector," Proc. IRE, Vol. 43, October 1955, pp 1354-1361.

18. K. Bullington, W. Inkster, A. Durkee, "Results of Propagation Tests at 505 Mc and 4090 Mc on Beyond-the-Horizon Paths," Proc. IRE, October 1955, pp 1306-1316.

19. M. Siddiqui, and G. Weiss, "Families of Distributions for Hourly Median Power and Instantaneous Power of Received Radio Signals," Journal of Research, NBS, Nov-Dec 1963, pp 753-762.

20. E. N. Gilbert, "Capacity of a Burst-Noise Channel," BSTJ, Vol 39, Sept. 1960, pp 1253-1265.

21. A. Alexander, R. Gryb, D. Nast, "Capabilities of the Telephone Network for Data Transmission," BSTJ, Vol 39, May 1960, pp 431-476.

22. S. Berkovits, E. Cohen, N. Zierler, "A Model for Digital Error Distributions," Proc. First IEEE Annual Communications Convention, Boulder, Colorado, June 1965, pp 103-111.

23. J. M. Berger and B. Mandelbrot, "A New Model for Error Clustering in Telephone Circuits," IBM Journal of Research and Development, Vol. 7., No. 3., July 1963, pp 224-236.

24. S. Sussman, "Analysis of the Pareto Model for Error Statistics on Telephone Circuits," IEEE Trans. on Communications Systems, Vol. CS-11, June 1963, pp 213-221.

25. F. E. Froehlich, D. Hirsch, H. R. Rudy, "High-Speed Digital Data Transmission over the Telstar Satellite," Proc. IRE, April 1963, p 609.

26. W. Bailey, "Data Transmission via Syncom Satellite," NSA Report S-155648, January 1964.

27. E. J. Baghdady, Lectures on Communication System Theory, McGraw-Hill, Book Co., 1961.

28. W. Ebenau, and K. Stotesbury, "Measuring Techniques for Establishing a Circuit for Medium-Speed Data Transmission," Technical Memorandum TM-86, Shape Technical Centre, Sept. 1964.

29. J. R. Tucker, J. Duffy, R. Salva, "Engineering Evaluation of Digital Data Transmission Capabilities of Army Communications Equipment," Mil-E-Con Proceedings, 1963, pp 205-222.

30. E. J. Hoffman, "Error Statistics on the Hawaiian Cable," Lincoln Laboratory Report 25G-0006, Feb. 19, 1960.

31. J. Bartow, "Testing of Digital Data Transmission Systems," U. S. Army Electronics Command, Technical Report ECOM-2558, Feb 1965.

32. A. B. Fontaine, "Error Statistics and Coding for Digital Data Transmission over Telephone and Teletype Circuits," Lincoln Laboratory Report 25G-14, Jan. 21, 1963.

33. A. B. Fontaine, and R. G. Gallager, "Error Statistics and Coding for Binary Transmission over Telephone Circuits, Proc. IRE, June 1961, pp 1059-1065.

34. E. J. Hoffman, "Error Statistics on Dual Data Circuits from Kingston, New York to Cape Canaveral, Florida," Lincoln Laboratory Report 25G-0005, Oct. 20, 1960.

35. E. J. Hoffman and P. L. Grant, "Error Statistics on a Long-Haul Looped Telephone Circuit Utilizing the Milgo Data System," Lincoln Laboratory Report 25G-0010, June 17, 1960.

36. R. Townsend, and R. Watts, "Effectiveness of Error Control in Data Communications over the Switched Telephone Network," BSTJ, Nov 1964, pp 2611-2638.

37. Motorola, Inc., "Data Transmission Investigation," Final Report, Contract DA 36-039-SC-90728, Sept. 11, 1963 to Oct. 10, 1964.

38. "Interstation Data Transmission Study Final Report," Philco Corporation, WDL-TR2017, Feb. 28, 1963.

39. Armour Reserach Foundation, "Methods of Comparison and Evaluation of Communication Systems," Vol. I, March 31, 1960. RADC-TR-60-97A.

40. A. Crawford and W. Jakes, "Selective Fading of Microwaves," BSTJ, Jan 1952, pp 68-90.

41. "Study and Application of Adaptive Communication Tecnhiques in the Defense Communication System," RCA Report CR-64-419-12, Sept. 5, 1964.

42. E. Sunde, "Digital Troposcatter Transmission and Modulation Theory," BSTJ, Jan 1964, pp 143-214.

43. Joint Technical Advisory Committee, "Radio Transmission by Ionospheric and Tropospheric Scatter," Proc. IRE, Jan 1960, pp 4-44.

44. J. C. Hancock, et al., "Advanced Communication Theory Techniques," ASD-TDR-63-186, March 1963.

45. A. Durkee, D. Metcalfe, W. H. Tidd, "White Alice System-Design and Performance," IRE Trans. on Communications Systems, Dec 1959, pp 272-277.

46. F. E. Wilson and W. A. Runge, "Data Transmission Tests on Tropospheric Beyond-the-Horizon Radio Systems," IRE Trans. on Communications Systems, March 1960, pp 40-43.

47. P. L. Grant, "Transmission Testing on Alaskan Forward-Propagation Tropospheric Scatter Circuit," Lincoln Laboratory Report 25G-9, May 29, 1962.

48. B. Goldberg, "HF Radio Data Transmission," IRE Trans. on Communication Systems, March 1961, pp 21-28.

49. A. R. Schmidt, "A Frequency Stepping Scheme for Overcoming the Disastrous Effects of Multipath Distortion on High Frequency FSK Communication Circuits," IRE Trans. on Communication Systems, March 1960, pp 44-47.

50. Stelma, Inc., "Development of Digital Data Terminal AN/GYC-1 for Short-Range HF Transmission Circuits," Report No. 2, Contract DA 28-043-AMC-00313(E), Oct 31, 1964 - Jan 30, 1965.

51. Armour Research Foundation, "Methods of Comparison and Evaluation of Communcation Systems," Vol. II, RADC-TR-60-97B, March 31, 1960.

52. A. J. Viterbi, "Phase-Coherent Communication over the Continuous Gaussian Channel," Digital Communications with Space Applications, edited by S. W. Golomb, Prentice-Hall, 1964.

53. K. Bullington, "Characteristics of Beyond-the-Horizon Radio Transmission," Proc. IRE, Oct 1955, pp 1175-1180.

54. D. G. Brennan, M. L. Phillips, "Phase and Altitude Variability in Medium Frequency Ionospheric Transmissions," Tech. Report No. 93, Lincoln Laboratory, 1957.

55. T. Hagfors and B. Landmark, "Simultaneous Variation of Amplitude and Phase of Gaussian Noise, with Applications to Ionospheric Forward Scatter Signals," Proc. IEE, Nov. 1958, pp 55f.

56. H. B. Voelcker, "Phase-Shift Keying in Fading Channels," Proc. IEE, Jan 1960, pp 31-38.

57. K. W. Otten, "Design of Reliable Long-Distance Air to Ground Communication Systems Intended for Operation Under Severe Multipath Disturbances," IRE Trans on Aerospace and Navigational Electronics, June 1962, pp 67-78.

58. C. R. Cahn, "Performance of Digital Phase-Modulation Communication Systems," IRE Trans on Communication Systems, May 1959, pp 3-6.

59. R. J. Benice and A. H. Frey, Jr., "Error Control through Coding, "Vol. VI, "The Design of Error Control Systems," RADC-TDR-64-149, July 1964.

60. A. H. Nuttall, "Error Probabilities for Uncorrelated M-ary Signals Under Phase-Coherent and Phase-Incoherent Reception," IRE Trans. on Information Theory, July 1962, pp 305-314.

61. P. Bello and B. D. Nelin, "Prediction Diversity Combining with Selectively Fading Channels," IRE Trans. on Communication Systems, March 1962, pp 32-42.

62. A. B. Crawford, D. C. Hogg, and W. H. Kummer, "Studies in Tropospheric Propagation Beyond the Horizon," BSTJ, Sept 1959, pp 1067-1178.

63. B. Reiffen, W. Schmidt, H. Yudkin, "The Design of An Error-Free Data Transmission System for Telephone Circuits," Communication and Electronics, July 1961, pp 224-231.

64. H. C. A. van Duuren, "Typendruktelegrafic over Radioverbindinger TOR," Tijdschrift van het Nederlands Radiogenootschap, March 1961.

65. K. G. Budden, "A Reciprocity Theorem on the Propagation of Radio Waves via the Ionosphere," Proc. Cambridge Phil. Soc. 50, Part IV, 1954, pp 604f.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) IBM, Federal Systems Division Rockville, Maryland	2a REPORT SECURITY CLASSIFICATION Unclass 2b GROUP N/A

3 REPORT TITLE

Adaptive Modulation and Error Control Techniques

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report

5 AUTHOR(S) (Last name, first name, initial)

Benice, R. J.

6 REPORT DATE May 1966	7a TOTAL NO OF PAGES 228	7b NO OF REFS 63

8a CONTRACT OR GRANT NO AF30(602)-3603 b PROJECT NO 4519 c Task No. 451902 d	9a ORIGINATOR'S REPORT NUMBER(S) 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-66-169

10 AVAILABILITY/LIMITATION NOTICES: This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of RADC (EMLI), GAFB, N.Y. 13440.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY Rome Air Development Center (EMCRS) Griffiss Air Force Base, New York 13440

13 ABSTRACT

This report contains the results of an investigation design to develop, analyze, and evaluate techniques to improve the quality of digital data transmission by adapting the modulation or error control parameters to varying channel conditions. Specific systems for varying the bit rate and the alphabet size are postulated and evaluated analytically in realistic channel environments. Variable block length and variable redundancy error control techniques are described and analyzed in a computer simulation of real channels. In addition, the operation and performance of a laboratory prototype of an adaptive error correction decoder that does not require feedback is described.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Adaptive Modulation, Adaptive Coding Communication Channel Models Information Theory						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED

Security Classification

SUPPLEMENTARY

INFORMATION

4S4 188

UNCLASSIFIED

RADC-TR-66-169 May 1966

ERRATA – August 1966

The following corrections are applicable to RADC-TR-66-169 entitled "Adaptive Modulation and Error Control Techniques," Unclassified Report, dated May 1966:

Page iii, line 1 change design to read designed.

Page iii, line 10 omit the second "decoder that".

Page 3-23 last line change between errors to read between bursts of errors.

Page 3-33 line 3 change $m2 - \pi$ to read $m2\pi + \pi$

Page 3-36 line 6 change $\frac{D}{3}$ to read $\frac{2}{3}(\frac{D}{3})$

Page 3-41 line 5 change order of $\frac{1}{B}$ to read order of $\frac{1}{B_s}$

Page 4-8 equation (4-5) change $F(\theta)f(\theta)\,d\theta - \sum_i C_i$ to read $\int F(\theta)f(\theta)\,d\theta - \sum_i C_i$

Page 4-55 Paragraph 4 line 1 change "modffication to read modification"

Page 4-87 figure 4-23 row 1 column 2 change D_{m+2} to D_{m+1}.

Page 4-88 line 3 change correcting bursts to read correcting virtually all bursts.

Page 5-6 equation (5-11) change $R_N \log_2 M_2\, \gamma(Z)\, dZ$ to read $R_N \log_2 M_N \gamma(Z)\, dZ$.

Page 5-7 equation (5-15) add + ... after $R_2 \log_2 M_2\, \gamma(Z)\, dZ$ and change

$R_n \log_2 M_n\, \gamma(Z)\, dZ$ to read $R_N \log_2 M_N\, \gamma(Z)\, dZ$.

Page 6-18 Figure 6-3 change Single Shot 120 Ms to Single Shot 120 μs.

Page 6-21 lines 5 and 6 change correcting 2800-bit bursts with a guard space of 4700 bits to read correcting virtually all 2800-bit bursts with an average guard space of 4700 bits.

Communications Research Branch
Rome Air Development Center
Research and Technology Division
Air Force Systems Command
Griffiss Air Force Base, New York